Jukka Suhonen

Designs for the Quality of Service Support in Low-Energy Wireless Sensor Network Protocols

Tampere 2012
Jukka Suhonen

Designs for the Quality of Service Support in Low-Energy Wireless Sensor Network Protocols

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB111, at Tampere University of Technology, on the 24th of August 2012, at 12 noon.
The research work for this Thesis was carried out in the Department of Computer Systems at Tampere University of Technology during the years 2005-2012.

I would like to express my gratitude to my supervisor Professor Marko Hännikäinen for the guidance, support, and motivation during the research. I am also grateful to Professor Timo D. Hämäläinen for his guidance and support. I would like to thank Associate Professor Evgeny Osipov from Luleå University of Technology and Professor Riku Jäntti from Aalto University for reviewing and providing comments for this Thesis. Also, I would like to thank Professor Timo T. Hämäläinen from University of Jyväskylä and Associate Professor Evgeny Osipov for agreeing to act as opponents in the defense.

I would like to express thanks to my co-authors Dr. Mikko Kohvakka and Dr. Mauri Kuorilehto for their valuable input in my research work, and also to Ville Kaseva, M.Sc., Teemu Laukkarinen, M.Sc., Lasse Määttä, M.Sc., Jani Arvola, M.Sc., and other members of the TUTWSN team for their work that made this Thesis possible. I am especially grateful to Markku Hänninen, M.Sc. for his valuable work in the extensive testing of the protocols presented in this Thesis. Also, I would like to thank Dr. Olli Lehtoranta for both research related and not so related discussions.

This work was financially supported by Graduate School in Electronics, Telecommunications and Automation (GETA).

Finally, I would like to express my gratitude to my family for their support and encouragement.

_Tampere, July 2012_

_Jukka Suhonen_
# TABLE OF CONTENTS

Abstract ................................................................. i
Preface ................................................................. iii
Table of Contents ..................................................... v
List of Publications ................................................... ix
List of Abbreviations .................................................. xi

1. Introduction ......................................................... 1
   1.1 WSN Design Characteristics .................................. 1
   1.2 Embedded WSN Platforms ................................. 3
   1.3 Quality of Service in WSNs .............................. 5
   1.4 Scope, Objectives, and Methods of Research ............ 6
   1.5 Results and Contributions ............................... 7
   1.6 Thesis Outline ........................................... 8

2. Applications and Standards ........................................ 9
   2.1 Applications ............................................... 9
   2.2 Wireless Communication Technologies .................... 10
   2.3 WSN Communication Standards ............................ 11
   2.4 Technology Integration via Internet of Things .......... 14
   2.5 Conclusion on Standards .................................. 15

   3.1 QoS Standards in Computer Networks .................... 17
   3.2 Performance Analysis .................................... 18
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3 Network Diagnostics</td>
<td>19</td>
</tr>
<tr>
<td>3.3.1 Passive monitoring</td>
<td>20</td>
</tr>
<tr>
<td>3.3.2 Deployment Support Networks</td>
<td>20</td>
</tr>
<tr>
<td>3.3.3 In-Network Diagnostics</td>
<td>21</td>
</tr>
<tr>
<td>3.4 QoS Protocols</td>
<td>21</td>
</tr>
<tr>
<td>3.5 Medium Access Control Layer</td>
<td>22</td>
</tr>
<tr>
<td>3.5.1 Channel Access Techniques</td>
<td>22</td>
</tr>
<tr>
<td>3.5.2 Low Duty Cycling</td>
<td>24</td>
</tr>
<tr>
<td>3.5.3 QoS Support in MAC</td>
<td>26</td>
</tr>
<tr>
<td>3.6 Routing Layer</td>
<td>27</td>
</tr>
<tr>
<td>3.6.1 Node-centric routing</td>
<td>28</td>
</tr>
<tr>
<td>3.6.2 Location-based Routing</td>
<td>28</td>
</tr>
<tr>
<td>3.6.3 Multipath Routing</td>
<td>28</td>
</tr>
<tr>
<td>3.6.4 Data-centric Routing</td>
<td>29</td>
</tr>
<tr>
<td>3.6.5 Cost-based Routing</td>
<td>30</td>
</tr>
<tr>
<td>3.6.6 QoS-aware Routing Proposals</td>
<td>30</td>
</tr>
<tr>
<td>3.7 Transport Layer</td>
<td>32</td>
</tr>
<tr>
<td>3.8 Summary</td>
<td>33</td>
</tr>
<tr>
<td>4. TUTWSN Platform and Deployments</td>
<td>35</td>
</tr>
<tr>
<td>4.1 Medium Access Control</td>
<td>35</td>
</tr>
<tr>
<td>4.2 Routing Protocol</td>
<td>36</td>
</tr>
<tr>
<td>4.3 Hardware Prototypes</td>
<td>37</td>
</tr>
<tr>
<td>4.4 Deployments</td>
<td>40</td>
</tr>
<tr>
<td>4.4.1 Outdoor Environmental Monitoring in Rural Area</td>
<td>41</td>
</tr>
<tr>
<td>4.4.2 Outdoor Environmental Monitoring in Suburban Area</td>
<td>42</td>
</tr>
<tr>
<td>4.4.3 Indoor Deployment at TUT Campus</td>
<td>42</td>
</tr>
</tbody>
</table>
### Table of Contents

5. *QoS Analysis for WSNs* ........................................... 45

5.1 QoS Metrics ..................................................... 45

5.1.1 Latency ..................................................... 46

5.1.2 Throughput .................................................. 46

5.1.3 Reliability and Availability ................................ 46

5.1.4 Network and Node Lifetime ................................ 48

5.1.5 Node Density, Count, and Communication Range ........ 48

5.1.6 Mobility ..................................................... 49

5.1.7 Security ..................................................... 49

5.2 Usage of QoS Profiles ........................................... 50

5.2.1 ZigBee Network Example .................................... 51

5.3 QoS Metrics in TUTWSN ......................................... 52

5.3.1 Application layer ............................................ 53

5.3.2 Routing layer ................................................ 54

5.3.3 MAC layer .................................................... 54

5.3.4 Physical layer ................................................ 55


6.1 QoS Schemes for WSN MACs .................................... 57

6.1.1 QoS Support Layer for WMNs ............................... 57

6.1.2 Dynamic Capacity Allocation ............................... 59

6.2 QoS Routing Cost Algorithm .................................... 62

6.2.1 Minimum Cost Routing ..................................... 62

6.2.2 Cost Algorithm .............................................. 63

6.2.3 Cost Functions for TDMA-based MAC ..................... 65

6.2.4 Simulation Results .......................................... 66

6.3 Cross-layer Design .............................................. 69
7.  Network Diagnostics ............................. 73
    7.1  Diagnostics Architecture ..................... 73
    7.2  Embedded Self-diagnostics ..................... 74
        7.2.1  Node information ......................... 74
        7.2.2  Network and node events ................... 74
        7.2.3  MCU and transceiver activity ............... 75
        7.2.4  Route and routing latency ................... 76
        7.2.5  Cluster and link traffic ................... 76
        7.2.6  Network topology .......................... 77
        7.2.7  Software errors .......................... 77
    7.3  Diagnosed QoS Metrics ....................... 78
    7.4  Performance Analysis Tool ..................... 78
    7.5  QoS Analysis on an Outdoor Network ............ 80
    7.6  Performance Comparison of Indoor and Outdoor Deployments ... 83
8.  Summary of Publications ........................... 85
9.  Conclusions ....................................... 87
This Thesis consists of an introductory part and the following publications. In the introductory part, the publications are referred to as [P1]...[P6]. The publications are sorted in chronological order based on the publication date.


<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6LoWPAN</td>
<td>IPv6 over Low power Wireless Personal Area Networks</td>
</tr>
<tr>
<td>ADC</td>
<td>Analog-to-Digital Converter</td>
</tr>
<tr>
<td>AODV</td>
<td>Ad-hoc On-demand Distance Vector routing</td>
</tr>
<tr>
<td>APS</td>
<td>Application Support</td>
</tr>
<tr>
<td>ATM</td>
<td>Asynchronous Transfer Mode</td>
</tr>
<tr>
<td>BO</td>
<td>Beacon Order</td>
</tr>
<tr>
<td>CAP</td>
<td>Contention Access Period</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant Bit Rate</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CFP</td>
<td>Contention-Free Period</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>CoAP</td>
<td>Constrained Application Protocol</td>
</tr>
<tr>
<td>COTS</td>
<td>Commercial Off-The-Shelf</td>
</tr>
<tr>
<td>CoS</td>
<td>Class of Service</td>
</tr>
<tr>
<td>CSMA</td>
<td>Carrier Sense Multiple Access</td>
</tr>
<tr>
<td>CSMA/CA</td>
<td>Carrier Sense Multiple Access with Collision Avoidance</td>
</tr>
<tr>
<td>CTS</td>
<td>Clear-To-Send</td>
</tr>
<tr>
<td>CW</td>
<td>Contention Window</td>
</tr>
<tr>
<td>DCF</td>
<td>Distributed Coordination Function</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>DCF</td>
<td>Distributed Coordination Function</td>
</tr>
<tr>
<td>DiffServ</td>
<td>Differentiated Services</td>
</tr>
<tr>
<td>DoS</td>
<td>Denial-of-Service</td>
</tr>
<tr>
<td>DSN</td>
<td>Deployment Support Network</td>
</tr>
<tr>
<td>DSSS</td>
<td>Direct Sequence Spread Spectrum</td>
</tr>
<tr>
<td>EEPROM</td>
<td>Electrically Erasable Programmable Read-Only Memory</td>
</tr>
<tr>
<td>FDMA</td>
<td>Frequency Division Multiple Access</td>
</tr>
<tr>
<td>GoS</td>
<td>Grade of Service</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio Service</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>HAL</td>
<td>Hardware Abstraction Layer</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>IEC</td>
<td>International Electrotechnical Commission</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IntServ</td>
<td>Integrated Services</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LPL</td>
<td>Low Power Listening</td>
</tr>
<tr>
<td>LR-WPAN</td>
<td>Low-Rate Wireless Personal Area Network</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MCU</td>
<td>Micro-Controller Unit</td>
</tr>
<tr>
<td>MIB</td>
<td>Management Information Base</td>
</tr>
<tr>
<td>MIPS</td>
<td>Million Instructions Per Second</td>
</tr>
<tr>
<td>NS2</td>
<td>Network Simulator 2</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NWK</td>
<td>Network</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>PAN</td>
<td>Personal Area Network</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>PHY</td>
<td>Physical layer</td>
</tr>
<tr>
<td>PIR</td>
<td>Passive Infra-Red</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RFID</td>
<td>Radio Frequency Identification</td>
</tr>
<tr>
<td>RPL</td>
<td>Routing Protocol for Low-Power and Lossy Networks</td>
</tr>
<tr>
<td>RSSI</td>
<td>Received Signal Strength Indicator</td>
</tr>
<tr>
<td>RSVP</td>
<td>Resource Reservation Protocol</td>
</tr>
<tr>
<td>RTS</td>
<td>Request To Send</td>
</tr>
<tr>
<td>RX</td>
<td>Reception</td>
</tr>
<tr>
<td>SNMP</td>
<td>Simple Network Management Protocol</td>
</tr>
<tr>
<td>SO</td>
<td>Superframe Order</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>SRAM</td>
<td>Static Random Access Memory</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>TOS</td>
<td>Type of Service</td>
</tr>
<tr>
<td>TRP</td>
<td>Transport</td>
</tr>
<tr>
<td>TUTWSN</td>
<td>Tampere University of Technology Wireless Sensor Network</td>
</tr>
<tr>
<td>TX</td>
<td>Transmission</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UI</td>
<td>User Interface</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>URI</td>
<td>Universal Resource Identifier</td>
</tr>
<tr>
<td>WIA-PA</td>
<td>Wireless network for Industrial Automation – Process Automation</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
<tr>
<td>WMAN</td>
<td>Wireless Metropolitan Area Network</td>
</tr>
<tr>
<td>WMN</td>
<td>Wireless Mesh Network</td>
</tr>
<tr>
<td>WPAN</td>
<td>Wireless Personal Area Network</td>
</tr>
<tr>
<td>WSN</td>
<td>Wireless Sensor Network</td>
</tr>
<tr>
<td>WWAN</td>
<td>Wireless Wide Area Network</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

In the vision of future networking, devices co-operate for intelligent decision making thus allowing unobtrusive operation without human interaction [2]. This enables a vast amount of applications in various areas such as military surveillance [127], security and asset management [14], environment monitoring [177], health care [103], building and home automation [165], and industrial control [57]. Generally, a sensor network refers to any set of interconnected sensor devices, including computers, home appliances, and mobile phones. This Thesis concentrates on low-energy Wireless Sensor Networks (WSNs), where tiny, unobtrusive sensor nodes gather information from surrounding environment, detect and classify events, and control actuators according to the detected events [169]. Compared to other wireless technologies, WSNs are characterized by low cost and ultra low energy [145]. This allows the deployment of even thousands of potentially disposable devices that can have a battery powered lifetime of years or operate on energy gathered from their environment. However, as a trade-off, the low energy WSNs have limited computation, communication, memory, and energy resources. Thus, the challenge is to ensure an adequate level of service.

This Thesis focuses on Quality of Service (QoS) in low energy WSNs. This Thesis concentrates to the performance at the network traffic level. As such, this Thesis considers some metrics that are typically referred to as constraints in the protocol design but still evaluate the performance from the application point-of-view. The main research problem is defining and implementing QoS with constrained energy budget, processing power, communication bandwidth, and data and program memories. The problem is approached via protocol designs and scheduling algorithms.

1.1 WSN Design Characteristics

A WSN consists of nodes that are deployed in the vicinity of an inspected phenomenon [4] as depicted in Fig. 1. A network typically contains one or more sink nodes that collect sensor values from other nodes. Instead of sending raw data to the
A WSN is typically deployed to perform a specific task, e.g. environmental monitoring, target tracking, or intruder alerting. As a result, WSNs are often data-centric in the sense that messages are not send to individual nodes but to geographical locations or regions based on the data content \[118\]. The application specific approach allows reducing communication overhead via data aggregation, and in-network processing and decision making \[32\].

Low energy nodes are typically battery powered but can also scavenge energy from their environment \[23\]. As replacing batteries may not be feasible due to large network size and energy scavenging does not typically produce enough power for continuous transceiver activity \[134\], network lifetime should be maximized via energy-efficient protocol designs.

Network density may be high as several nodes are located in close proximity. Still, a WSN may operate in large geographic areas and contain a vast amount of sensor nodes. This has several implications. First, a network technology must be scalable to ensure that performance does not degrade even on large networks. Second, to reduce deployment and maintenance effort, a network must be autonomous and self-configurable. Third, transmitting data directly to a target node is not feasible as the

![Fig. 1. An example WSN scenario presenting data collection in a multihop topology.](image-url)
required (free space) transmission energy is proportional to the square of the distance [5] with obstacles further reducing the communication range [179]. Thus, covering large geographical areas implies multihop routing.

Table 1.1 summarizes the typical WSN characteristics and their implications to the protocol and hardware designs. A WSN may not share every characteristic, e.g. the scalability is not a primary concern on few nodes deployments.

### 1.2 Embedded WSN Platforms

A WSN platform comprises tightly coupled hardware and software. It determines the performance and energy resources that are available for applications, thus having a significant effect on the level of service.

WSN platform consists of four basic units [4] that are necessary for sensing, processing values, and delivering measurements to the locations where they can be exploited:

- **Sensing unit** measures physical phenomena via sensors, controls actuators, and convert measurements to digital values with Analog-to-Digital Converter (ADC).

- **Computing unit** that typically comprises a microprocessor to execute instructions, persistent program memory for application code, temporary data mem-

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Benefits</th>
<th>Design implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application specific</td>
<td>Optimizations based on known traffic</td>
<td>Aggregation and in-network processing and decision making</td>
</tr>
<tr>
<td>Long network lifetime</td>
<td>Battery powered nodes with the lifetime of years</td>
<td>Low power platforms and energy-efficient protocols</td>
</tr>
<tr>
<td>Small physical node size</td>
<td>Unobtrusive, even wearable nodes</td>
<td>Resource constraints limiting the computational and memory complexity of application and protocols. Battery size restrictions limiting available energy</td>
</tr>
<tr>
<td>Low cost</td>
<td>Economic feasibility of a large number of potentially disposable nodes</td>
<td></td>
</tr>
<tr>
<td>Ease of maintenance</td>
<td>Unplanned and low-labor deployment</td>
<td>Autonomous and self-configurable operation</td>
</tr>
<tr>
<td>Scalability</td>
<td>Large network size and high density for accurate sensing</td>
<td>Scalable protocol design and multihop operation</td>
</tr>
</tbody>
</table>

*Table 1. Typical WSN characteristics and their implications to the protocol and hardware designs.*
1. Introduction

ory such as Static Random Access Memory (SRAM), and persistent data memory such as Electrically Erasable Programmable Read-Only Memory (EEPROM) or flash.

- Communication unit that connects node to network via wireless transceiver. The transceiver typically uses Radio Frequency (RF) technology as it does not have the line-of-sight requirement of infrared and ultrasound.

- Power unit provides energy for other components e.g. via energy scavenging, batteries, or mains power.

The computing unit has the most diverse functionality as it manages collaboration between nodes and carries out sensing tasks. To ease development, a node may use an Operating System (OS) that manages memory, provides Hardware Abstraction Layer (HAL) for sensors and other hardware resources, and allows interaction between application tasks [92, 168]. The communication between nodes is managed by a protocol stack that contains physical, Medium Access Control (MAC), routing, and transport layers. The physical layer exchanges bits over a physical link between nodes. MAC manages neighbor discovery, establishes wireless links, and exchanges frames with neighbors by receiving and transmitting on wireless channel [84], while routing enables end-to-end communications over multiple hops [118]. The transport protocol ensures reliable end-to-end transmission of packets and congestion control [192]. Instead of accessing the network stack directly, an application may use a middleware layer that provides providing application frameworks and interfaces e.g. for collaboration between nodes, security, localization, and runtime configuration on heterogeneous hardware [144]. Based on hardware capabilities, WSN nodes may be classified to high performance and low energy platforms [64]. The high performance platforms have computing and memory capabilities that are close to Personal Computers (PCs) whereas low energy platforms aim at low cost and long lifetime on batteries. A network may be heterogeneous and comprise both kind of nodes, as nodes can be specialized in certain tasks.

This Thesis concentrates on the low energy platforms that allow the deployment of large scale and long term sensor networks [57]. Due to the limitations in the manufacturing techniques, low energy, low cost, and small size can be realized only with a resource constrained hardware [128]. A low energy WSN node has typically only few Million Instructions Per Seconds (MIPSs) processing power, 32-128 kB program memory, and 2-8 kB data memory [84].
1.3 Quality of Service in WSNs

QoS has various meanings depending on context. Generally, it describes whether a service satisfies user expectations and includes traffic performance, security level and quality of technical support [76]. ITU-T makes a difference between Grade of Service (GoS) and QoS in its E.600 and E.800 recommendations. GoS is a subset of QoS that concentrates on measuring the traffic performance [75]. In this Thesis, QoS is considered only from GoS point of view, and both terms are used interchangeably.

In communication networks, QoS is usually understood as a set of performance requirements to be met for transferring a data flow [33]. These requirements are defined and measured with a set of quantifiable attributes referred to as QoS metrics [150]. In legacy computer networks, QoS is commonly expressed with throughput, delay, jitter (variation of transfer delays), and error rate metrics [56].

In this Thesis, a protocol that implements a control to differentiate at least one QoS metric is referred to as a QoS protocol. Thus, a QoS protocol adapts its operation to meet the QoS demands. In practice, QoS is realized in communication protocols that give either soft (relative) or hard (absolute) service level guarantees.

The importance of QoS is emphasized in wireless networks that suffer from unreliable communications, link quality, link breaks, and limited communication capacity. In WSNs, these issues are especially evident due to the unplanned deployment that causes low quality links, and energy depletion that leads to node failures. While QoS has been researched in traditional computer networks, the existing QoS protocols are too complex for the resource constrained sensor nodes [193] and do not consider energy that is important for WSNs. This necessitates the design of new QoS protocols. The state of the art research on sensor QoS has concentrated on single metrics such as energy or latency.

The potential use cases for WSNs vary significantly and have different requirements. A simple measurement network that collects periodic samples tolerates high latencies and low reliability, since the sensed physical phenomenon changes slowly and few packet misses can be tolerated. Alert messages, such as fire or intruder detection, can tolerate small, few second delays but high reliability is critical. Control traffic that is used for interaction between users and devices necessitates low latency and high reliability. While the throughput requirements for all of these applications are low, high capacity WSNs may also be used e.g. for multimedia streaming that require high bandwidth [3, 113]. As a single network may comprise traffic from different classes, QoS support is needed to fulfill the service level requirements.
Table 2 shows an example use classes in industrial WSNs based on the patterns of intended use, specified by the ISA standard organization in its ISA100.11a standard [74]. In process industry, latency and reliability are often critical but monitoring applications can tolerate delays while human triggered control actions (open loop) and automatic control actions (closed loop) have strict timing and reliability requirements. Traffic that triggers emergency actions must always be delivered with very low latency and high reliability. In the context of ISA specifications, the scope of this Thesis are the low energy protocols that are suitable for classes 3-5. The other classes are meant for automated control with very high reliability requirements and latencies in order of milliseconds.

To ensure that the network performance meets the desired QoS, network diagnostics is required both in protocol testing and practical deployments. Although some of the issues can be eliminated with a careful deployment, a practical network might have software failures, logical errors in protocols and algorithms, and node failures due to energy depletion or hardware failures. Identifying problems in a large scale deployment is particularly challenging as problems may reflect to several parts of the network. This necessitates diagnostics to detect and identify the performance issues.

### 1.4 Scope, Objectives, and Methods of Research

The scope of this research consists of QoS definition and protocols for low energy, resource constrained WSNs. QoS is considered on MAC, routing, and transport layers as presented in Fig. 2. Sensing applications are covered based on their service requirements. Application specific algorithms, data aggregation [29, 46], sensing [172], and hardware designs are outside the scope of this Thesis.

The first objective of this Thesis is to define QoS for low energy WSNs to enable

<table>
<thead>
<tr>
<th>Category</th>
<th>Class</th>
<th>Description</th>
<th>Criticality of latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety</td>
<td>0</td>
<td>Emergency action</td>
<td>Always critical</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>Closed loop regulatory control</td>
<td>Often critical</td>
</tr>
<tr>
<td>Control</td>
<td>2</td>
<td>Closed loop supervisory control</td>
<td>Usually non-critical</td>
</tr>
<tr>
<td>Control</td>
<td>3</td>
<td>Open loop control</td>
<td>Non-critical</td>
</tr>
<tr>
<td>Monitoring</td>
<td>4</td>
<td>Alerting</td>
<td>Non-critical</td>
</tr>
<tr>
<td>Monitoring</td>
<td>5</td>
<td>Data logging</td>
<td>Non-critical</td>
</tr>
</tbody>
</table>
1.5 Results and Contributions

The main results of this Thesis are

- A survey of existing QoS communication protocols and standards for low energy WSNs [P1-P6],
- Definition of metrics that allow assessing QoS quantitatively [P4],
- QoS support layer for Wireless Mesh Networks (WMNs) [P5],
- QoS control algorithm for WSN MACs [P2,P6],

quantitative performance comparisons between different networks. The second objective is to design communication protocols that realize the defined QoS in practice. The third objective is to develop methods to measure and manage QoS in WSNs, thus allowing verification that the network performance met the user expectations.

The research started by identifying the QoS issues and requirements with a literature review and examining of the requirements of typical sensor applications. These results were used as a basis for defining the WSN QoS definition and protocol designs for QoS. The protocols were verified with simulations on Network Simulator 2 (NS2) tool, prototype implementations in Tampere University of Technology Wireless Sensor Network (TUTWSN) [91], and real-world deployment studies. TUTWSN is a WSN technology developed in the Department of Computer Systems at Tampere University of Technology (TUT) for low data rate monitoring applications. The TUTWSN platform was used to verify the practical feasibility of the results of this Thesis. As an exception, the protocol presented in [P5] was tested in IEEE 802.11 Wireless Local Area Network (WLAN) [68] environment. Finally, embedded self-diagnostics were designed and utilized to analyze the performance in deployments.

Fig. 2. The scope of this Thesis is on protocol designs at MAC and routing layers.
• Energy-efficient QoS routing protocol [P1],
• WSN self-diagnostics defining collected performance data on a sensor node and how the data is transmitted to the gateway for further analysis [P3], and
• Diagnostics tool to analyze the collected diagnostics information [P3].

1.6 Thesis Outline

The Thesis consists of an introductory part and 6 publications [P1]-[P6]. The introductory part motivates the work, presents technical background, and summarizes the results. The results are presented in the publications.

The rest of the introductory part is organized as follows. WSN application space, WSN related standards, and the research background on QoS protocols are provided in Chapters 2 and 3. Chapter 4 presents the TUTWSN platform that was used in the implementations and presents the deployments that were used to verify the results of this Thesis. The rest of the Chapters describe the results of this Thesis: Chapter 5 defines QoS metrics for WSNs, Chapter 6 composes the research results on QoS enabled WSN protocol design, and Chapter 7 presents sensor self-diagnostics framework and diagnostics tools for measuring and analyzing WSN QoS. The publications included in this Thesis are summarized in Chapter 8. Chapter 9 concludes the Thesis.
2. APPLICATIONS AND STANDARDS

This chapter presents the main requirements for sensing applications, covers the current key communication standards of the area, and discusses their applicability to WSNs.

2.1 Applications

While the application domain for WSNs is diverse, the applications can be classified with few basic characteristics. In general, a WSN may execute one or more of the following application tasks [4, 80]:

- **Data logging**: A node measures certain physical phenomenon e.g. temperature, humidity, or luminance. The measurement may be triggered periodically or when a change is detected.

- **Event detection**: A node monitors and detects an event of interest, e.g. motion or a sensor reading that exceeds certain limits.

- **Object classification**: A node processes sensor values to identify the type of object or event, possibly combining values from several sensors. For example, the network might determine the type of moving object (animal, human, vehicle, etc.).

- **Object tracking**: Sensor information is used to trace the movement path of a mobile object based on location, direction, and speed estimates.

- **Control**: A node controls actuators, such as light switches or valves, based on direct commands from an user or an automation system, or by making independently decisions based on measured sensor values.

The tasks listed are complementary to each other, and a task does not need to be active all the time. Many tasks require collaborative operation between nodes, e.g. combining values from several sensor nodes to give more accurate sensor value or
object classification. As an example, a surveillance network may perform continuous data logging, while the collected data is used as a basis for event detection. After an event (e.g. motion) is detected, the network classifies the moving object. Object tracking could be activated only when an unauthorized object is detected.

2.2 Wireless Communication Technologies

Wireless communication technologies are categorized based on their typical coverage and application domains [49, 65, 112]. The link range, data rate, mobility, and power requirements of the technologies are presented in Fig. 3. The values are not definite but illustrate the differences between the technologies. In the figure, RF communications is assumed as it is most widely used and does not have inherent limitations such as line-of-sight requirement in infrared.

Wireless Wide Area Network (WWAN) covers a large geographical area and consists of telecommunications networks such as Global System for Mobile Communications (GSM) and satellite communications. In telephone networks, broadband data is supported with packet-switched data services such as General Packet Radio Service (GPRS) or Universal Mobile Telecommunications System (UMTS). Mobility requirements are critical, as uninterrupted service is expected even when a user is traveling on high-speed rail (200+ km/h) [125].

Wireless Metropolitan Area Network (WMAN) covers geographic area or region that is smaller than WWAN but larger than WLAN. An examples of WMANs is IEEE 802.16 (WiMAX) [71]. Both WWAN and WMAN use highly asymmetric devices, as simpler end devices connect to base stations. As such, these networks are intended for single hop uses where the wireless access is used to connect to the Internet or global telephone network [35]. Wireless multihop support is rare and typically limited to base stations.

WLAN spans a relatively small area, such as building or a group of buildings. IEEE 802.11 [68] is the dominant WLAN technology. It was originally targeted to access a wired Local Area Network (LAN) with wireless interface but has been since extended to support mesh networking in 802.11s extension. IEEE 802.11 is widely utilized for network access in public buildings and enterprises, and sharing Internet in homes.

Wireless Personal Area Network (WPAN) is a short distance network for interconnecting devices centered around an individual person including watches, headsets, mobile phones, audio/video equipment, and laptops. Bluetooth [15] and IEEE 802.15 standard family [69, 70] are the most widely used WPAN technologies. WPANs have
2.3. WSN Communication Standards

Fig. 3. Properties of wireless communication technologies.

varying energy and throughput requirements as the use cases range from low power data exchange with portable devices to high data rate home entertainment and multimedia transfers.

WSN shares most properties with WPANs and may utilize similar technologies. For example, IEEE 802.15.4 low-rate WPAN standard [70] is used as a basis for many WSN communication standards. However, a WSN is designed for multiple users, has usually more devices, and often emphasizes lifetime.

2.3 WSN Communication Standards

Standards promote interoperability between products from different manufacturers. Table 3 lists standards and industry specifications suitable for WSNs. The support for Physical layer (PHY), MAC, Network (NWK), and Transport (TRP) denotes that the technology defines the layer in question. Application Support (APS) defines application profiles that detail services, message formats, and methods required to access applications.

IEEE 802.15.4 Low-Rate Wireless Personal Area Networks (LR-WPANs) uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) for channel access and supports real-time applications via guaranteed time slots. A network comprises three types of devices: a Personal Area Network (PAN) coordinator, coordinators, and end devices. Coordinators are more complex but can route data while the end devices can be realized with simpler hardware. A network may operate in two modes. In a non-beacon enabled operating mode, the coordinators listen to the channel continuously therefore necessitating mains power. In a beacon enabled mode, coordinators transmit periodic beacon frames that are used for synchronization. A beacon identifies PAN and describe the structure of the following superframe. Beacons allow low duty
### Table 3. Key WSN communication standards.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Freq. band (MHz)</th>
<th>Data rate (kbps)</th>
<th>Protocol layers</th>
<th>PHY</th>
<th>MAC</th>
<th>NWK</th>
<th>TRP</th>
<th>APS</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE 802.15.4</td>
<td>868</td>
<td>20</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>IEEE 802.15.4</td>
<td>915</td>
<td>40</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>IEEE 802.15.4</td>
<td>2400</td>
<td>250</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>ZigBee</td>
<td>-</td>
<td>-</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Bluetooth Low Energy</td>
<td>2400</td>
<td>1000</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Z-Wave</td>
<td>865</td>
<td>40</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Z-Wave</td>
<td>915</td>
<td>40</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>MiWi</td>
<td>2400</td>
<td>250</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>ANT/ANT+</td>
<td>2400</td>
<td>1000</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>WirelessHART</td>
<td>2400</td>
<td>250</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>ISA100.11a</td>
<td>2400</td>
<td>250</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>WIA-PA</td>
<td>2400</td>
<td>250</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>ONE NET</td>
<td>868/38.4</td>
<td></td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>ONE NET</td>
<td>915</td>
<td>230</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>DASH7</td>
<td>433</td>
<td>27.8</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>IEEE 1902.1 RuBee</td>
<td>0.131</td>
<td>1.2</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

- ● defined in standard, ○ not defined, ● reuse of IEEE 802.15.4 PHY

Cycle operation where nodes wake up to receive beacons and participate in superframe, but remain in low power sleep state most of the time.

ZigBee technology [215] defines network and application layers on top of the IEEE 802.15.4. A device referred to as a ZigBee coordinator controls the network. The coordinator is the central node in the star topology, the root of the tree in the tree topology, and can be located anywhere in the peer-to-peer topology. ZigBee defines a wide range of application profiles targeted at home and building automation, remote controls, and health care.

MiWi [48] specified by Microchip Technology Inc. is a simplified version of the ZigBee. It uses IEEE 802.15.4 non-beacon enabled mode and supports small networks up to 1024 nodes.

Z-Wave [210] is targeted at building automation and entertainment electronics. A typical Z-Wave network contains a mixture of AC powered and battery powered nodes. The lifetime of routing nodes is very limited, as they listen continuously to the channel. The maximum number of nodes in a network is 232. Supported network topologies are star and mesh. Z-Wave has been developed by over 120 companies.
2.3. WSN Communication Standards

including Zensys, Intel and Cisco.

WirelessHART [73] and ISA100.11a [74] are targeted at process industry applications where process measurement and control applications have stringent requirements for end-to-end communication delay, reliability, and security. The standards have similar operating principle and the convergence of the standards is planned in ISA100.12 [158]. Both standards build on top of the IEEE 802.15.4 physical layer and utilize a Time Division Multiple Access (TDMA) MAC that employs network wide time synchronization, channel hopping, channel blacklisting. A centralized network manager is responsible for route updates and communication scheduling for entire network. However, as the centralized control of TDMA schedules limits the network size and the tolerance of a WSN node against network dynamics, the usability of the standards is limited to relatively static networks.

Wireless network for Industrial Automation – Process Automation (WIA-PA) is originally a Chinese specification for industrial automation but is also approved as an international standard by International Electrotechnical Commission (IEC) [96]. WIA-PA uses IEEE 802.15.4 physical and MAC layers. The standard specifies how the guaranteed time slots of IEEE 802.15.4 are allocated, defines adaptive frequency hopping, and allows aggregating several short packets into one packet to reduce overhead. Compared to WirelessHART and ISA100.11, WIA-PA is more adaptable to varying traffic loads but does not have as good real-time guarantees due to the limited amount of contention-free in IEEE 802.15.4 [214].

Bluetooth Low Energy (BLE) is an extension to the Bluetooth technology [15] aimed at low energy wireless devices. Devices advertise their presence with periodic beacons, while listening to the channel briefly for incoming connection or data requests after each advertisement. Data is exchanged with attribute/value pairs. Advertisements can also contain data and connections are established fast (less than 3 ms), therefore avoiding the need to stay in connected state and enabling devices to save energy in standby states.

ANT [43] defined by Dynastream Innovations Inc. is used e.g. by Suunto and Garmin in their performance monitoring products. ANT is based on virtual channels that are defined by operating frequency and message rate parameters. Due to TDMA based communications, several channels may operate on same physical frequency. Master nodes always receive, while slaves transmit when new data is provided. Complex topologies can be formed as each node may act as a master and a slave on different channels. ANT+ extension includes profiles defining data formats and channel parameters.
ONE-NET [182] is an open-source WSN specification comprising MAC and routing protocol designs and example hardware schematics. It operates on 868/915 MHz with the data rate of 38.4-230 kbps. ONE-NET supports low duty cycling for battery powered devices but routing nodes must keep their transceivers active thus necessitating mains power.

DASH7 [153] technology based on ISO 18000-7 standard is targeted at very low rate data applications. Its main cited benefit stems from the 433 MHz operating frequency, which provides longer communication ranges and less crowded wireless channel than the typical 2.4 GHz frequency band [121]. DASH7 has the nominal communication range of 250 m at 0 dBm transmission power level, compared to 75 m of ZigBee and 10 m of Bluetooth (High Rate variant) [121].

IEEE 1902.1 (RuBee) [67] fills the gap between WSN and Radio Frequency Identification (RFID) technologies. Unlike other listed technologies, signal does not include electric field component but uses magnetic dipole antennas. Thus, signal is unaffected by water and metals either enhance or do not affect the signal. RuBee nodes, referred to as tags, can be very simple identity tags or use 4-bit MCU, 0.5 kB-2 kB SRAM, optional sensors, signal processing firmware, displays and buttons [124]. The nominal data rate is small, 1.2 kbps, limiting the applicability of RuBee.

2.4 Technology Integration via Internet of Things

Internet of Things (IoT) is a new paradigm that aims to integrate heterogeneous communication technologies to the part of global network infrastructure in a way that they can be used seamlessly with each other [218]. This enables co-operation and interaction between a variety of things or objects, e.g., RFID tags, sensors, actuators, mobile phones, and computers [8]. In practice, this means making individual sensors and actuators addressable anywhere from the Internet either by using Internet Protocol (IP) for end-to-end communications, or via gateway adaptation by converting messages to technology specific formats and mapping network’s internal addresses to the global IP addresses.

Internet Engineering Task Force (IETF) has defined IPv6 over Low power Wireless Personal Area Networks (6LoWPAN) standard to describe how IP is used in IEEE 802.15.4 networks. It addresses IP routing on mesh networks and defines methods to allow transmitting large sized IP packets in bandwidth constrained environments. Other features addressed in 6LoWPAN include network autoconfiguration and multicast emulation. 6LoWPAN is also being adapted for Bluetooth Low Energy
2.5. Conclusion on Standards

The header compression and unicast and multicast methods defined for IEEE 802.15.4 are reused in the Bluetooth Low Energy adaptation, while the Bluetooth address mapping to IP addresses is defined.

Constrained Application Protocol (CoAP) is an interface protocol especially targeted for constrained networks and machine-to-machine applications such as smart energy and building automation [161]. The protocol operates over User Datagram Protocol (UDP) and is designed according to the REST architecture. As such, the protocol is easy to map to Hypertext Transfer Protocol (HTTP) and Universal Resource Identifiers (URIs), therefore enabling web integrated sensors. Messages can be cached to improve performance, e.g. at the gateway to avoid requesting data directly from sensor network. Other features supported by CoAP are multicast support and congestion control.

2.5 Conclusion on Standards

WSNs are being deployed for a wide range of uses, each with varying QoS requirements. This demands QoS protocols that allow adjusting service level based on application demands. While a wide range of WSN standards have been introduced, none of the standards cover the entire WSN application space. Instead, each standard optimizes its operation for a certain use case.

Current WSN standards have been mostly lacking controllable QoS support. ZigBee uses link reliability in its routing but does not otherwise consider QoS. IEEE 802.15.4 supports bandwidth reservations for real-time traffic and throughput guarantees. However, the mechanism is requires explicit reservation handshaking, making it mainly applicable for Constant Bit Rate (CBR) traffic. Also, an upper layer reservation protocol is required to trigger the reservation process. ISA100.11a, WirelessHART, and ANT enable fine grained QoS by allowing network wide control of channel access times. This enables delay and throughput guarantees for end-to-end flows, but requires prior knowledge of network traffic and disallows dynamic traffic. Thus, the applicability of the protocols is limited to relatively static networks or simple devices, where traffic consists of predetermined polling.
2. Applications and Standards
3. RELATED RESEARCH ON QoS METRICS AND PROTOCOLS

This chapter discusses the related work on QoS methods and protocols relevant to this Thesis. First, existing QoS standards and their suitability for WSNs is discussed. Then, models and frameworks that measure and manage QoS are presented. Finally, this chapter surveys QoS communication protocols proposed in the literature.

3.1 QoS Standards in Computer Networks

QoS has been extensively studied in wireless LANs and wired computer networks. For example, IP [132] and Asynchronous Transfer Mode (ATM) [178] provide extensive QoS support ranging from best-effort service to guaranteed service. The QoS models in IP can be divided in the following categories: best-effort, relative priority marking, service marking, label switching, Differentiated Services (DiffServ) [117], and Integrated Services (IntServ) [18]. The best-effort service is the simplest and means that QoS is not specifically addressed.

The other service models provide a variable degree of QoS. The IP header contains a precedence field for providing relative priority marking and a Type of Service (TOS) / Differentiated Services (DS) field for providing service marking [117]. Relative priority marking and service marking describe the desired service within the IP header of a packet. The priority marks the importance of the packet (e.g. delay and drop priority). The service marking allows selecting a routing path that prefers either delay, throughput, reliability, or (monetary) cost.

Label switching [146], DiffServ and IntServ operate on traffic aggregates instead of marking a single packet. Label switching is used within a single network to route data along a specific path. In DiffServ, the traffic entering the network is classified and each class is assigned with different behavior. This approach to QoS is referred to as Class of Service (CoS). IntServ provides service guarantees by defining two types of services: guaranteed service and guaranteed load service [18]. The guaranteed service uses reservations to enable end-to-end QoS and to guarantee the wanted
throughput together with maximum delays [162].

QoS has been addressed also on cellular and WMANs. However, the protocols used in these networks differ greatly from WSNs as they have only one wireless hop located between an end device and a base station. A backbone network is usually wired and based on circuit switching, IP, or ATM.

The QoS standards in computer networks are computationally complex, utilize extensive routing tables requiring large data memory, or impose high communication overhead due to extensive messaging. Thus, they are not applicable to the resource constrained WSNs [4, 107, 213].

3.2 Performance Analysis

QoS need to be defined to allow comparing the user requirements to the realized performance. While several studies have evaluated the QoS related challenges and unique problems of WSNs [3, 27, 193, 206], only few of the published articles try to define WSN QoS. Dietrich and Dressler [37] aim to formalize network lifetime definition by mapping node availability, sensor coverage, and network connectivity metrics to the lifetime. Qiang et al. [133] define a QoS evaluation model which maps application layer parameters to network layer parameters with fuzzy logic. The application layer parameters comprise data accuracy, network lifetime, response time, and event detection probability. The network layer parameters consist of energy-efficiency, packet delay, throughput, and reliability. However, these works examine QoS only partially, and mainly from the perspective of sensor coverage. Network layer is considered only with the traditional QoS metrics.

The research on network performance measurements concentrates mainly on detecting misconfigured nodes [106], compromised nodes [38], software assertions and remote debugging [87, 198], and collecting sensor readings and detecting anomalies in the sensor data [45, 217]. While the detected anomalies can trigger a distributed self management e.g. to compensate the fault by increasing sensing threshold on other sensors [148], the failure is often only reported to the gateway[207]. Generic purpose UI tools offer a framework for visualizing sensor networks but do not consider the actual methods to analyze the data [208].

Only few papers consider actual QoS performance measurements. Ringwald and Römer [141] list possible performance problems and their causes on WSN deployments but do not specify any methods to measure or detect the issues. Software architecture that considers energy, neighbor, and link quality diagnostics is presented
3.3. Network Diagnostics

In computer networks, Simple Network Management Protocol (SNMP) has become the de facto network management and monitoring protocol [154]. SNMP is an application layer protocol that accesses virtual information storage, referred to as a Management Information Base (MIB), located at the target device. As a MIB contains device and protocol specific information, hundreds of specifications have been defined, both by IETF and by other organizations as manufacturer specific extensions. Kim et al. [82] have proposed MIB for 6LoWPAN, comprising device address and role (coordinator, router, non-router), device capability information (e.g., can device be a coordinator), type of power source, and the enumeration of routes and known neighbors. Other diagnostics information, such as reliabilities, were not considered.

Research effort has been made to improve the suitability of SNMP for constrained devices by reducing traffic overhead [31, 154]. While the SNMP design is considered relatively light weight [154], its memory requirements can still be prohibitive for resource constrained nodes. In [93], it was found that a full implementation of SNMP on an AVR Raven platform with 6LoWPAN required 30.5 kB program memory and 1 kB data memory, which correspond to 24% and 7% of total, respectively. A limited implementation with SNMPv1 and without authentication/privacy options reduced the memory requirements to 8.6 kB (7%) program and 0.47 kB (4%) data memory. Still, SNMP can be used e.g., with a gateway acting as proxy/adapter that converts in-network WSN diagnostics to SNMP [154].

As determining the level of performance and identifying potential network problems are important both for end-users and developers, many current network technologies
support at least basic in-network diagnostics. For example, ZigBee allows querying node’s approximate energy level (near empty, half, full), and node’s neighbors and link qualities [216]. Due to their tight reliability requirements, WirelessHART and ISA100 standards have extensive diagnostics support. The supported features comprise remaining lifetime estimate, neighbor and traffic information, and the average latency from gateway to device.

The related research proposals on network diagnostics can be categorized to passive monitoring, deployment support networks, and in-network diagnostics. These are discussed in the following sections.

3.3.1 Passive monitoring

Passive monitoring tools are typically portable devices that listen to the WSN traffic. They can be used without explicit planning and do not cause any overhead to the monitored network. The simplest passive monitoring tools act as packet sniffers, while more advanced functionality includes bandwidth usage and connectivity analysis. A complete network view can be formed by combining data from several packet analyzers, thus allowing to reconstruct the network topology, determine bandwidth usage and routing paths, make connectivity analysis, and to identify hot-spot nodes. For example, Chen et al. in [26] log packets to Flash memory until the packet sniffers are manually retrieved later. An offline software merges and analyzes the packet traces. The multi-sniffer approach proposed in [199] allows visualizing the otherwise complex behavior of a WSN with a graphical view. The proposed system is also able to replay the recorded network activities at different speeds.

3.3.2 Deployment Support Networks

Deployment Support Network (DSN) refers to a separate diagnostics network that is installed alongside the actual sensor network [13]. As such, the DSNs are typically short-lived and removed once the network operation is verified. A DSN node has typically two radios: one for overhearing WSN traffic, and a second for forming the support network to forward the overheard packets to a gateway. LiveNet proposed in [26] uses wired Ethernet for collecting data. This ensures that wireless interference does not affect diagnostics collection but also increases the network deployment effort. [142] and [42] define wireless DSNs that use Bluetooth scatternet with up to 100 mW transmission power. The high performance radio reduces the lifetime of the proposed DSNs to few weeks with two AA-size batteries. In general, due to doubled
hardware and increased costs DSNs are best suited for protocol testing and development.

3.3.3 In-Network Diagnostics

In in-network diagnostics, nodes collect diagnostics information concerning their operation and pass it to the gateway using the same communication protocols and radios than the sensor data. The approach has two distinct benefits. First, additional diagnostics equipment is not needed which makes collecting the diagnostics throughout the lifetime of the network feasible. Second, nodes can include information about their internal operation and decisions making. The main drawback is that the diagnostics information consumes bandwidth in an already resource constrained network.

A diagnostics method presented in [101] piggybacks status information to data packets. The method marks a packet with a forwarding node identifier if the packet is received out-of-sequence. This allows detecting the existence and location problems but does not reveal reasons for problems (e.g. interference or bad link). Visibility metrics introduced in [190] considers the cost (e.g. in terms of bandwidth or energy) to collect diagnostics from a network. The metric constructs tree-based decision graph that contains potential problems in the network. The visibility cost is calculated by assigning a probability and information collection cost to each problem. However, the paper does not specific any metrics.

3.4 QoS Protocols

The communication protocols presented in this section are categorized based on the layers for which they are primarily targeted at. However, the distinction between the layers is not always clear because many WSN communication protocols implement features that traditionally belong to several layers [107, 167]. This enables low complexity protocol designs that meet the resource constraints [6]. Also, it improves overall network performance as cross-layer information allows more optimal forwarding decisions [213].

The related work is limited to the protocols that consider network QoS. Application specific QoS, such as sensor coverage [172], sensor accuracy [173], exposure $^1$, and measurement errors [27] are outside the scope of this Thesis. Furthermore, as the

$^1$ Exposure denotes locating sensors in such way that effects from obstacles are minimized
number of protocol proposals is very large\(^2\), the protocols in this Chapter are meant to be representative examples of discussed techniques instead of exhaustively naming each protocol.

### 3.5 Medium Access Control Layer

A MAC layer manages transmissions and receptions on a shared wireless medium, therefore having a significant impact on performance and energy consumption [84]. This section describes the typical WSN MAC design principles and their effect to QoS. The protocols are classified based on their channel access technique and the use of duty cycling as presented in Fig. 4.

#### 3.5.1 Channel Access Techniques

MAC protocols can be categorized into contention and contention-free protocols. Alternatively, these are also referred to as random and scheduled protocols [59]. In contention-based protocols, bandwidth is divided among nodes on-demand basis. The method achieves low latencies and relatively good bandwidth utilization on lightly loaded networks or when the number of contending nodes is low [36]. When a network is loaded by multiple nodes, a collision avoidance scheme is required to prevent significant performance degradation. In WSNs, contention protocols typically utilize CSMA/CA [184] which checks channel activity prior to a transmission and defers a transmission for a random backoff interval, referred to as Contention Window (CW), to avoid collisions [36, 39, 175, 204]. The backoff time is a compromise

\(^2\) As an example, a search on IEEE Xplore digital library concerning routing protocols alone, with wireless, sensor, network, and routing in the publication title, produced 1504 publications between years 2001-2011.
between latency, bandwidth utilization efficiency, and collision probability.

Typical problems in the contention-based protocols comprise the hidden node (hidden terminal) problems [185] and idle listening. Hidden node problem can be prevented with an additional signaling such as Request To Send (RTS)/Clear-To-Send (CTS) mechanism or a combination of carrier sensing and control packets. However, these increase communication overhead [89]. The idle listening is a result of unknown transmission times, which necessitates a receiver to sense channel continuously for incoming packets [187, 203].

In contention-free schemes, transmissions are arranged for collision free channel access. The typical contention-free schemes are polling, TDMA, Frequency Division Multiple Access (FDMA), and Code Division Multiple Access (CDMA) [89]. Predetermined channel access minimizes idle listening, avoids collisions, and enables accurate QoS control, as transmission times and capacity can be assigned deterministically. However, the drawbacks are synchronization and slot assignment overhead. In traditional TDMA systems, Transmission (TX)/Reception (RX) slots are assigned by a central manager which reduces scalability. Therefore, many WSN proposals use distributed methods where nodes exchange known reservation information within two-hop neighborhood [25, 61, 94, 140].

Another problem in the contention-free protocols is determining the correct amount of reservations. As a monitored physical phenomenon may generate traffic bursts when an event triggers, slot usage increases momentarily and unpredictably. Furthermore, traffic varies even with CBR sources as varying channel conditions cause link breaks, packet errors, and retransmissions. As a result, capacity is either over or under reserved. Unused capacity is wasted and consumes energy due to unnecessarily reception, while too low reserved capacity increases transfer delays and may cause packet losses [79]. For these reasons, a pure contention-free scheme is mainly applicable for static or centrally controlled networks [89]. However, a contention-free protocol can react to traffic bursts by reserving only a part of the slots, while using other slots dynamically on-demand. For example, in Y-MAC [83] a node operates initially on a certain base channel but uses additional channels for traffic bursts. After a successful reception, a node switches another channel and listens to the next time slot. Still, the reservation problem for the base channel slots remain in Y-MAC.

Hybrid approaches aim to combine the flexibility of contention-based protocols to the energy-efficiency and reliability of the contention-free techniques. IEEE 802.11 and IEEE 802.15.4 protocols support both schemes. However, the contention-free access is usually used only for guaranteed throughput. The Contention-Free Period (CFP)
slot reservations are constant and changes has to be negotiated between communicating nodes, making the approach taken in these protocols inflexible. Traffic adaptive medium access protocol (TRAMA) [136] alternates between random access and scheduled periods. Nodes may join a network only during the random access period. Then, nodes exchange information in two-hop neighborhood about their intended receivers and construct transmission time schedule based on this information. Thus, the protocol avoids assigning time slots to nodes without traffic therefore minimizing idle listening. Z-MAC [139] combines CSMA and TDMA based approaches. Time is divided into communications slots where each slot may be assigned to a certain node (owner). CSMA is used in each slot but owners use a shorter backoff time, thus giving them an earlier chance to transmit. Other nodes may steal the slot if it is not used by its owner. Thus, Z-MAC uses TDMA scheme as a hint to enhance contention resolution. The scheme allows robustness to various slot assignment failures and topology changes.

### 3.5.2 Low Duty Cycling

Due to the requirement for long term deployments and battery powered operation, most of the proposed sensor MAC protocols concentrate on lifetime maximization [4, 89, 156]. In the wireless networks, transceiver consumes most energy [52, 189]. Thus, a common goal for energy-efficient MAC protocols is reducing the transceiver activity by minimizing idle listening, collisions, and protocol overhead [84].

Several WSN protocols utilize low duty cycle operation, in which duty cycle (transceiver activity) is adjusted to the network traffic therefore minimizing the idle listening. Although duty cycling decreases energy usage, it increases forwarding latency due to sleeping delay: a node must wait until the next active time before a packet can be forwarded [204].

Duty cycling may use either synchronized or unsynchronized approach as illustrated in Fig. 5. In the synchronized method, a node maintains a periodic sleep schedule consisting of active and idle periods [134]. The synchronization is commonly realized by transmitting beacon frames at the beginning of an active period. The repeated period is referred to as an access cycle or a superframe. A node may need to wake up multiple times during an access cycle to forward data if its neighbors use different schedules. In a typical approach, a node receives data during its active period. The active period can be realized with either contention or contention-free channel access technique. During the idle period, a node forwards data to its neighbors (participates another node’s active period) or saves energy by sleeping.
IEEE 802.15.4 LR-WPAN supports a synchronized low duty cycle approach where coordinators have temporally non-overlapping active periods within the interference range. In contrast, S-MAC [203, 204] and its derivatives, including T-MAC [187], nanoMAC [58], DSMAC [99], RMAC [39], and DW-MAC [175], aim to use a common sleep schedule (overlapping active periods) between nodes. The approach reduces control messaging but is applicable only for contention-based channel access. All packets, including beacons, are transmitted with Carrier Sense Multiple Access (CSMA) utilizing RTS/CTS mechanism. This allows relaxing synchronization requirements, and it reduces overhead as there is no need to transmit beacons every access cycle. T-MAC [187] improves S-MAC by adjusting active period length dynamically based on traffic requirements. An active period ends when traffic is not received within a certain time interval.

NanoMAC [58] adds a support for block acknowledgments. DSMAC [99] keeps active period length constant but scales access cycle length (sleep time) according to traffic. RMAC [39] improves S-MAC by sending control frame via multiple hops that assigns reception schedules so that routing latency is minimized. DW-MAC [175] adds support for contention-free channel access by using sleep periods for communication: frames transmitted in data period reserve proportional portion of sleep period.

Unsynchronized low duty cycle protocols use typically Low Power Listening (LPL) mechanism. In LPL, nodes poll channel asynchronously to test for incoming traffic instead of transmitting regular beacons. Transmissions are preceded with a preamble that acts as a wake-up signal. For correct operation, the preamble must be at least as long as poll interval. A node detecting the preamble listens to the channel until a packet is received or a timeout occurs. As the preamble is often longer than the actual transmission, a node may experience significant delay as it must wait until other

![Fig. 5. Beacon synchronized and unsynchronized low duty cycle channel access techniques. Node A forwards a data frame to node B.](image-url)
transmissions are complete [176]. For example, in Fig. 5 (right) node A misses one of its periodic polls while transmitting the preamble. Thus, LPL MACs are mainly suitable for light traffic loads [171]. LPL is used in IEEE 1902.1 and DASH7 standards.

The proposed enhancements to the basic LPL scheme aim to optimize the preamble length [131, 205]. WiseMAC [44] attempts to improve the efficiency by reducing the duration of preamble transmission with a fixed wakeup schedule and frequent communication between neighbors. X-MAC [20] transmits multiple short preambles with the address of intended receiver. Upon receiving a short preamble, the destination node sends an acknowledgment between the preambles which triggers transmission. SCP-MAC [205] uses LPL mechanism with synchronized channel polling. This reduces energy as only short preamble is needed. The synchronization is realized by broadcasting periodic synchronization frames.

Receiver Initiated MAC (RI-MAC) [176] reverses the reception and transmission phases in the LPL scheme. Instead of transmitting a preamble, a sender turns its transceiver on and listens until the receiver transmits a beacon frame. This triggers the transmission. A receiver acknowledges frame with another beacon thus extending the active period and allowing traffic bursts. Beacon interval is randomized around a set value to reduce collisions. The method improves throughput and reliability over other LPL schemes. However, depending on the traffic patterns, the energy-efficiency can be lower as typical low power WSN transceiver consumes more energy in reception state due to employed de-spreading and error correction techniques [22].

3.5.3 QoS Support in MAC

MAC layer QoS concentrates on reliability, energy, and latency. Reliability is mainly ensured by controlling the amount of retransmissions [163]. Clustered operation increases network lifetime by dedicating energy consuming routing to cluster heads [90, 195]. Clustering is especially energy-efficient with synchronized protocols as only cluster heads need to transmit beacon frames and listen to the channel extensively. Member nodes listen only to the beacons unless they have data for the cluster head. To even energy consumption, the cluster head role may be rotated among the nodes belonging to a cluster [63].

Latency is controlled with priority based channel access and duty cycle scheduling approaches. The priority based channel access assumes CSMA/CA and assigns a high priority packet with a shorter contention window, thus allowing an earlier trans-
mission opportunity [102, 104, 200]. This approach was also taken in IEEE 802.11e which defines QoS for IEEE 802.11 WLAN networks. While the contention window length is typically determined by an application, other metrics can be used. Q-MAC [200] derives the number of transmitted hops, residual energy, and the proportional load of an output queue.

Duty cycle scheduling aims to adjust active periods in a manner that minimizes the sleeping delay. The active period of a next hop is located immediately after the active period of an forwarding node assuming that a frame from previous hop is received during own active [95, 105]. However, this kind of adjustment is mainly useful for optimizing routing delay toward one destination, e.g. when a network has one sink. This reduces the forwarding delay of packets that have traveled several hops and increases the importance of traffic from low energy nodes while avoiding overloading the wireless medium.

3.6 Routing Layer

As several alternative routes to a destination node may exist, each with different QoS properties, the route selection has a significant impact on QoS [60]. Furthermore, a single route with an optimal all-purpose QoS might not exist, thus necessitating the route selection based on application requirements. For example, one route might have minimum end-to-end latency while another route could be more reliable.

WSN routing protocols can be categorized based on their operation as node-centric, data-centric, location-based, multipath, or cost-based [5, 118]. The routing categories are presented in Fig. 6. These categories are not exclusive as a protocol can be both data-centric and query based.

Fig. 6. Categories of WSN routing protocols.
3. Related Research on QoS Metrics and Protocols

3.6.1 Node-centric routing

Node-centric approach is the traditional approach used in the computer networks in which nodes are addressed with globally unique identifiers [118, 147]. Node-centric protocols typically rely on routing tables containing an entry for each route identified by destination address and next hop node for the target. The routing table may be constructed proactively by discovering routes to all potential targets, but this increases memory requirements and would not be practical in large networks. Instead, the node-centric protocols designed for ad-hoc wireless networks, such as Dynamic Source Routing (DSR) [80], or Ad-hoc On-demand Distance Vector routing (AODV) [130], use a reactive approach in which routes are constructed only when needed. The drawback compared to the proactive approach is the route construction delay when sending first packets.

3.6.2 Location-based Routing

Location-based routing uses geographic location information to make routing decision. The approach is natural to WSNs, as sensor measurements usually relate to a specific location. A basic principle in the geographic routing is to select a next hop neighbor that is closer to the target node than a forwarding node [81, 119, 157]. Location-based routing is scalable as routing tables are not needed and a network can support a high degree of mobility. However, determining the position for each node can be problematic. The use of positioning chips such as Global Positioning System (GPS) increases the price and energy consumption, while manual configuration is not suitable for large scale networks.

3.6.3 Multipath Routing

In the multipath routing, a packet traverses from a source node to a target node via several paths [47, 51]. The main goal is to increase reliability, as a packet can be received via an alternative path even if the routing in some path fails. However, the multipath routing has a trade-off between the reliability and energy, as it increases network load and energy usage due to the extra transmissions. Flooding packet to every node in the network is the simplest case of multipath routing [30]. In flooding, each node forwards a new flood packet to all of its neighbors. To suppress duplicates, already received flood packets are not forwarded. Flooding is commonly used during
the setup phase of several WSN routing protocols, but is not used for routing as such because packets can easily congest network and thus decrease reliability.

Another approach to multipath routing is the constructing of several routing paths from a source to the destination, but using only one path at the time e.g., when a link breaks. For example, Localized Multiple Next-hop Routing (LMNR) extends AODV by storing several minimum hop paths to the destination [114]. The used route is selected based on local cost metric that aims at load balancing. Several alternative metrics for calculating the cost are proposed: size of IEEE 802.11 CW, outgoing buffer usage, packet leaving rate, or route table size and freshness of the routes.

3.6.4 Data-centric Routing

In data-centric routing, data is routed based on its content rather than using sender or receiver identifiers. As the data-centric routing is already content aware, data-aggregation can be naturally performed. Data centric routing may take interest, negotiation, or query approaches.

In the interest approach, a sink node request data from the network by sending a request describing the data it wants to every node in the network [5]. A node forwards the interest and directs its routing tree toward the sink. Then, nodes that fulfill the requirements as defined in the interest start transmitting data to the sink. Although the route construction is proactive, the interest based routing is scalable as the number of sinks (data consumers) is low compared to the number of nodes (data sources).

Negotiation protocols exchange messages before an actual data transmission takes place [88, 116, 137]. This saves energy, as a node can determine during the negotiation that the actual data is not needed. For negotiation protocols to be useful, the negotiation overhead and data descriptor sizes must be smaller than the actual data.

Query based routing protocols request a specific information from the network [17, 53]. A query might be expressed with a high level language such as Structured Query Language (SQL). For example, a query might request “average temperature around area x,y during the last hour”. The query can be routed via a random walk [160], flooded to the whole network [151], or directed at a certain region [100]. After the query has been resolved, the result is transmitted back to the source.
3.6.5 Cost-based Routing

In cost-based routing, each node is assigned with a cost value that is relative to the distance between a node and a sink [78, 209]. The cost may be calculated from an any metric, e.g. the number of hops, the required energy to forward a packet to the sink [1], or throughput [34, 98]. The benefit of the cost-based routing is that the knowledge of forwarding path states is not required: a node forwards its data by sending it to any neighbor that has lower cost. The drawback is that the routes must be created proactively. Also, although data to the sink is forwarded efficiently, another routing mechanism, such as flooding, must be used for data traveling in the other direction. However, the trade-off can be acceptable since most of the traffic is usually toward the sink.

IETF has defined Routing Protocol for Low-Power and Lossy Networks (RPL) as a routing protocol for constrained IPv6 networks [196]. RPL uses cost-based routing for node-to-sink (many-to-one) communications and node centric routing for sink-to-node traffic. Nodes periodically broadcast routing information to their neighbors which allows the detection of route changes. The routing cost is referred to as a rank, and it is calculated from one or more metrics with an objective function. A network may run multiple routing instances concurrently with different objective functions, e.g., separate instances for different sinks in the network. The instance is recognized from identifiers included in routed packets. The node centric routing in RPL has two alternative modes. In a non-storing mode, nodes advertise their parents to the root which then uses (DSR-like) source routing. In a storing mode, each node stores a routing table containing the reachable child nodes.

3.6.6 QoS-aware Routing Proposals

Several research proposals point out that finding the optimal route subject to multiple QoS constraints is NP-hard problem, necessitating very high communication overhead [194, 197]. As the overhead is unacceptable in the resource constrained WSNs, routing proposals typically aim towards approximate solutions [197].

Only few routing protocols consider route reliability [62, 86]. In ZigBee [215], a per link cost is calculated from a measured link reliability which improves throughput compared to traditional shortest hop-count routing protocols [86]. However, other important metrics, such as energy, are ignored. GRAB [202] uses controlled multi-path routing where a packet can be given a certain credit to allow extra transmissions and thus make a trade-off between energy and reliability. However, similar or higher
reliability and a smaller overhead could be achieved by using per-hop retransmissions [P4].

Many WSN routing proposals consider residual energy or forwarding energy [24, 212]. In [155], a next hop is selected randomly between routes having the same cost. If the energy remaining in a node is low, the node discourages others from routing through it by increasing its cost. The protocol in [159] uses similar cost approach but selects the forwarding node with a probability that depends on the energy metric of the route. The maximum lifetime routing [24] calculates the cost by combining the transmission and reception energy consumption with the residual energy of a node.

Some cost-based routing proposals target at minimizing maintenance energy by reducing the messaging overhead from the exchange of cost information. Minimum Cost Forwarding [201] uses a backoff algorithm to reduce the message overhead during the setup phase. The localized max-min remaining energy routing [10] minimizes the exchange of routing information between nodes by introducing an extra delay that is inversely proportional to the remaining energy. Thus, the route with the lowest delay has the lowest energy usage.

Sequential Assignment Routing (SAR) [166] forms a multipath tree rooted at a sink node. Next hop is selected by energy cost, QoS metric, and priority level of a packet. The QoS metric may be defined as required, for example, it may be based on link delay. On each hop, SAR calculates weighted cost from the link cost and priority level assigned for a packet. Thus, a higher priority packet result into a lower weighted cost and can traverse through nodes that have less energy but ensure higher QoS. The drawback of the SAR is that the changes in QoS metrics, energy, or topology require a recomputation of routes. SAR recovers from these changes by performing periodic updates initiated by the sink node.

Directed diffusion [72] is a data centric routing protocol that has motivated many proposals. It names data with attribute-value pairs and forwards data based on its contents rather than using sender or receiver identifiers. Initially, a sink requests data by injecting an interest into the network, where it gradually disseminates to each node. The sink refreshes its interest periodically to recover from unreliable interest propagation. When a node receives an interest, it establishes a gradient toward the sender node. Once the gradients have been established, directed diffusion offers energy-efficient node-to-sink data delivery [16]. Loops are prevented by maintaining a data cache that contains recently received items and data rates for each gradient. In self-stabilizing diffusion protocol [12] nodes can query cached interests from their neighbors. This allows better error recovery and reduces the need for refreshing inter-

CEDAR [164] uses core nodes located in a grid topology to form a QoS path. The destination node is found by flooding within the core. Then, a directed search is performed to find a QoS constrained path. The drawback with the CEDAR is that the network core may be broken at transient times due to network dynamics and the repair cost is high. CEDAR does not consider energy consumption.

Trajectory Based Forwarding (TBF) [119] is a location-based routing protocol that forwards a packet along a predefined curve. Energy consumption and forwarding delay can be controlled by choosing a next hop near the curve that has most energy left or is most forwarding link. Trajectory and Energy-Based Data Dissemination (TEDD) [55] uses the same concept as TBF but generates trajectories based on the global knowledge of remaining energies in sensor nodes (energy map). The problem with both TBF and TEDD is that the sender requires global knowledge of the network for QoS and avoiding obstacles.

SPEED [183] is a stateless non-deterministic location based routing protocol. It provides soft end-to-end guarantees that are proportional to the distance between source and destination nodes, thus determining certain delivery speed for a packet. MM-SPEED [47] extends SPEED by maintaining multiple speed information, therefore allowing several classes of service. It also increases reliability by using probabilistic multipath forwarding approach.

The IETF RPL supports the use of an arbitrary route cost metric that can be additive, multiplicative, minimum or maximum among the local link and node costs in a route. In addition, the specification allows using constraints related to a metric, thus preventing the selection of paths that do not meet a certain metric value. IETF has defined the remaining energy, hop count, throughput, latency, and link reliability metrics for RPL [188]. However, the effect of co-using these metrics is not specified.

### 3.7 Transport Layer

The main objectives of a transport protocol are ensuring end-to-end reliability, congestion control, and in order packet delivery [9]. The congestion control reduces transmission rate to prevent packet losses due to buffer overflows. Problems with the transport protocols used in the computer networks, such as Transmission Control Protocol (TCP), are high overhead, low fairness for nodes for nodes that are far
away from sink, and mistaking packet loss for congestion even when they are caused transmission errors on bad links [192].

In the proposed transport protocols for WSNs, the reliability is ensured either with end-to-end [77, 152] or hop-by-hop retransmissions [126, 126, 191]. The hop-by-hop transmission is generally more efficient in multihop WSNs as it requires less packet transmissions thus having better energy-efficiency [54]. Also, the hop-by-hop protocols can support sink to many nodes multicasting [126, 191].

As guaranteeing the delivery of all packets might not be necessary due to the redundancy of sensor data, few transport protocols consider event reliability [135, 152]. This way, only one packet related to an event, e.g. an intruder alert, is delivered while the network can save resources by dropping other similar packets.

One proposed transport protocol, referred to as asymmetric and reliable transport [181], considers energy by selecting a high energy node to report an event. As this only affects source nodes, energy balancing do not affect forwarding nodes. In general, although the proposed protocols aim toward reliability, they do not explicitly consider QoS.

3.8 Summary

The QoS definitions and performance evaluation methods proposed in the literature concentrate on only one or few aspects of QoS. This has motivated the design of QoS metrics, performance analysis methods, and WSN self-diagnostics in this Thesis.

Most of the related WSN protocols aim to optimize one aspect of QoS such as lifetime. As a drawback, different traffic types, such as non-critical measurements and critical alerts, have similar level of service. Few protocols support traffic differentiation but only with the regard of a specific metric (e.g. latency). As a result, the proposed protocols are optimal only for specific use cases.

This Thesis presents QoS designs at MAC and routing layers for traffic differentiation with multiple QoS metrics. Thus, the protocol designs enable heterogeneous WSN applications with varying QoS requirements to operate in the same network.
3. Related Research on QoS Metrics and Protocols
4. TUTWSN PLATFORM AND DEPLOYMENTS

This chapter describes the TUTWSN platform and the practical experiments used to verify the results of this Thesis. TUTWSN is a research platform targeted at low-energy applications requiring several years lifetime with AA-sized batteries. The platform contains research effort from several researchers, comprising hardware prototypes, network protocols, sensor operating systems, programming interfaces, embedded sensor applications, and services and applications outside the sensor network.

The author of this Thesis designed the queue control, error control, and bandwidth management methods for the MAC layer, routing protocol, and cross-layer QoS control for TUTWSN. These protocol designs for QoS are presented in detail in the following chapters. As the research of this Thesis concentrates on the embedded network protocol design, the other parts of the platform are discussed here with the depth that allows evaluating the results.

4.1 Medium Access Control

TUTWSN MAC [91] utilizes synchronized low duty cycle channel access. The active portion of a superframe comprises contention access and contention-free periods similarly to e.g. IEEE 802.15.4. TUTWSN MAC has few distinct design choices. First, network beacons are transmitted on a dedicated network channel for rapid and energy-efficient neighbor discovery as shown in Figure 7. This way, a node only listens to network channel instead of going through each cluster channel separately [84]. Second, Contention Access Period (CAP) is realized with slotted ALOHA to allow implementation on simple low cost transceivers that lack carrier sensing functionality.

The contention access and contention-free periods are divided into fixed length time slots. Each time slot is further divided into two subslots where the first subslot is for data transmission and the following subslot is for an acknowledgment. An example of message exchange during an active period of the superframe is presented in Figure 8.
In the example, a member node sends data to the cluster head during the CAP and later during the CFP. Another member node both sends and receives data during the CFP. All transmissions during the active period occur in the cluster channel.

In this Thesis, the MAC layer roles of nodes are defined as follows. A headnode is a node that acts as a cluster head in at least one cluster. A headnode can also join other clusters as a member node e.g. to forward data. A subnode is a node that acts only as a member node. Thus, a subnode does not send beacons or maintain CAP, which significantly reduces its energy consumption.

4.2 Routing Protocol

TUTWSN routing uses cost-based approach due to its simplicity and low overhead. As the cost-based approach exploits the notion that the number of data consumers (sinks) is usually much smaller than the number of data sources [171], the target applications of the routing are monitoring and target-tracking. The developed routing protocol extends the basic principle of cost-based routing with the support of multiple sinks, the use of interests to instruct data collection, localized route construction allowing fast recovery from broken links, and multiple QoS metrics.
The localized route construction is presented in Fig. 9. Initially, a node requests routes from its neighbors and selects the lowest cost route as its next hop. Nodes send route advertisements which allows their neighbors to detect new routes and recover from missed route requests/replies.

TUTWSN forms a multi-cluster tree topology as shown in Fig. 10. Headnodes forward data towards sinks while subnodes communicate with the nearest headnode. In the figure, two routing trees are constructed towards a sink node. The source node selects which tree to follow by comparing the QoS definition of the route against application requirements.

TUTWSN routing has features that are typically associated with the transport layer. End-to-end reliability is ensured by rerouting failed transmission on MAC layer via another link as described in publication [P4] while the congestion control is realized by considering traffic load in cost calculations. The routing protocol is presented in detail in [P1] while the QoS routing cost calculation is presented in Chapter 6.

### 4.3 Hardware Prototypes

The TUTWSN hardware prototypes are built with Commercial Off-The-Shelf (COTS) components. The platforms use Microchip PIC18LF8722 Micro-Controller Unit (MCU) with 128 kB flash program memory, 4 kB SRAM data memory, and 2 kB EEPROM memory. Apart from gateway devices that use mains power, the nodes are powered with two serially connected Lithium AA-size batteries with the total capac-

![fig9](image1)

**Fig. 9.** Route discovery in TUTWSN. a) Node A broadcasts a route request. b) Neighbors reply with an advertisement. c) Node A selects the lowest cost route. d) Nodes advertise periodically their routes. e) An advertisement reveals a better path.

![fig10](image2)

**Fig. 10.** Multi-cluster tree topology of TUTWSN.
ity of 3000 mAh.

TUTWSN has two platform variants, a long range (LR) variant that is targeted at outdoor deployments, and a low energy (LE) variant targeted at dense networks. The transceiver properties of these platforms are summarized in Table 4. In addition to the two variants, minor revisions of the platforms exist with varying sensors and an optional Ethernet connectivity. As these revisions do not affect network performance as such, they are not further detailed here. Circuit board antennas are used to minimize the physical size and cost of a node.

TUTWSN platform was selected because the author’s research was part of a larger research project with the focus of developing cross-layer designs with the joint optimization of hardware and network protocols. Currently, most widely used platforms in related literature include Berkeley motes and their variants [129]. Still, due to the application specific nature of WSNs, there is no de facto WSN platform. TUTWSN platform is compared to the motes and selected commercial platforms in Table 5. The platform list is not exhaustive, as technology specific products (e.g., application development kits for a specific standard) are not listed because the research of this Thesis required the full control of hardware resources. As a summary, the communication and computational resources of TUTWSN are in line with the related platforms and the TUTWSN prototypes act as examples of typical resource constrained and low energy WSN nodes [84].

Table 4. Transceiver properties of TUTWSN platforms.

<table>
<thead>
<tr>
<th>Property</th>
<th>Long range</th>
<th>Low energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transceiver</td>
<td>Nordic nRF905 [123]</td>
<td>Nordic nRF24L01 [122]</td>
</tr>
<tr>
<td>Frequency band</td>
<td>433 MHz</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Packet size, max</td>
<td>32 B</td>
<td>32 B</td>
</tr>
<tr>
<td>Nominal data rate</td>
<td>50 kbit/s</td>
<td>1 Mbit/s</td>
</tr>
<tr>
<td>Output power, max</td>
<td>10 dBm</td>
<td>0 dBm</td>
</tr>
<tr>
<td>Energy usage, RX mode</td>
<td>250 nJ/bit</td>
<td>11.3 nJ/bit</td>
</tr>
<tr>
<td>Energy usage, TX mode¹</td>
<td>600 nJ/bit</td>
<td>11.8 nJ/bit</td>
</tr>
<tr>
<td>Communication range²</td>
<td>500 m</td>
<td>180 m</td>
</tr>
<tr>
<td>Unique channels</td>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>

¹ Energy usage with the maximum transmission power.
² Range in open space with the optimum alignment of antennas.
Table 5. Comparison of TUTWSN to selected commercially available WSN platforms.

<table>
<thead>
<tr>
<th>Platform</th>
<th>MCU model</th>
<th>Clock (MHz)</th>
<th>Prgm (kB)</th>
<th>Data (kB)</th>
<th>EEPROM (kB)</th>
<th>Radio model</th>
<th>RF band (MHz)</th>
<th>Data rate (kb/s)</th>
<th>Size (mm²)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arduino Uno [7]</td>
<td>ATMega328</td>
<td>16</td>
<td>32</td>
<td>2</td>
<td>1</td>
<td>XBee ZB</td>
<td>2400</td>
<td>250</td>
<td>n/a</td>
<td>1</td>
</tr>
<tr>
<td>Arduino Mega [7]</td>
<td>ATMega2560</td>
<td>16</td>
<td>256</td>
<td>8</td>
<td>4</td>
<td>XBee ZB</td>
<td>2400</td>
<td>250</td>
<td>n/a</td>
<td>1</td>
</tr>
<tr>
<td>BTnode ver3 [219]</td>
<td>Atmega128L</td>
<td>8</td>
<td>128</td>
<td>64</td>
<td>4</td>
<td>CC1000</td>
<td>433/915</td>
<td>76.8</td>
<td>1919</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ZV4002</td>
<td>2400</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libelium Wasmote [97]</td>
<td>ATmega1281</td>
<td>8</td>
<td>128</td>
<td>8</td>
<td>4</td>
<td>XBee-802.15.4</td>
<td>2400</td>
<td>250</td>
<td>3749</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>XBee-900</td>
<td>900</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>XBee-868</td>
<td>868</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEMSIC MicaZ [109, 111]</td>
<td>ATMega128L</td>
<td>8</td>
<td>128</td>
<td>4</td>
<td>4</td>
<td>CC2420</td>
<td>2400</td>
<td>250</td>
<td>1856</td>
<td></td>
</tr>
<tr>
<td>MEMSIC IRIS [108, 111]</td>
<td>ATMega1281</td>
<td>8</td>
<td>128</td>
<td>8</td>
<td>4</td>
<td>RF230</td>
<td>2400</td>
<td>250</td>
<td>1856</td>
<td></td>
</tr>
<tr>
<td>MEMSIC TelosB [110]</td>
<td>MSP430</td>
<td>8</td>
<td>48</td>
<td>10</td>
<td>16</td>
<td>CC2420</td>
<td>2400</td>
<td>250</td>
<td>2015</td>
<td>2</td>
</tr>
<tr>
<td>Shimmer [138]</td>
<td>MSP430</td>
<td>8</td>
<td>48</td>
<td>10</td>
<td>0</td>
<td>CC2420</td>
<td>2400</td>
<td>250</td>
<td>1969</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RN-42</td>
<td>2400</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUTWSN LR</td>
<td>PIC18LF8722</td>
<td>8</td>
<td>128</td>
<td>4</td>
<td>2</td>
<td>nRF905</td>
<td>433</td>
<td>50</td>
<td>8100</td>
<td></td>
</tr>
<tr>
<td>TUTWSN LE</td>
<td>PIC18LF8722</td>
<td>8</td>
<td>128</td>
<td>4</td>
<td>2</td>
<td>nRF24L01</td>
<td>2400</td>
<td>1000</td>
<td>2800</td>
<td></td>
</tr>
</tbody>
</table>

1 XBee radio modules inter-changeable (only one attached at a time), 2 Platform has two integrated radios (WSN and Bluetooth)
Fig. 11 presents the examples of LE and LR nodes. The presented low energy node is equipped with temperature, humidity, and photo diode sensors, whereas the long range node has only temperature sensor.

4.4 Deployments

The protocol designs of this Thesis were verified in several real-life TUTWSN pilot studies in both outdoor and indoor environments. The pilots acted in the QoS research as a practical source of performance measurements. The pilots are summarized in Table 6. Two outdoor environmental monitoring and campus network deployments are detailed in the following sections.

![Fig. 11. TUTWSN low energy (left) and long range (right) nodes that were used to verify the results of this Thesis.](image)

<table>
<thead>
<tr>
<th>Deployment</th>
<th>Nodes</th>
<th>Duration</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sewer water level monitoring</td>
<td>25</td>
<td>2009-10</td>
<td>Long range</td>
</tr>
<tr>
<td>Chemical factory monitoring</td>
<td>62</td>
<td>2009</td>
<td>Low Energy</td>
</tr>
<tr>
<td>Green house monitoring</td>
<td>30</td>
<td>2009</td>
<td>Low Energy</td>
</tr>
<tr>
<td>Campus network for teaching</td>
<td>200</td>
<td>2008-</td>
<td>Low Energy</td>
</tr>
<tr>
<td>Residential monitoring⁴</td>
<td>180</td>
<td>2007-</td>
<td>Low Energy</td>
</tr>
<tr>
<td>Cargo monitoring and tracking⁴</td>
<td>50</td>
<td>2008-09</td>
<td>Low Energy</td>
</tr>
<tr>
<td>Building automation monitoring⁴</td>
<td>377</td>
<td>2008-</td>
<td>Low Energy</td>
</tr>
<tr>
<td>Outdoor environment monitoring⁴</td>
<td>100</td>
<td>2005-</td>
<td>Long range</td>
</tr>
<tr>
<td>Environment conditions in a cow house</td>
<td>30</td>
<td>2009-</td>
<td>Low Energy</td>
</tr>
</tbody>
</table>

¹ Several separate deployments, the amount of nodes is the total from these pilots.
4.4. Deployments

4.4.1 Outdoor Environmental Monitoring in Rural Area

An outdoor network using TUTWSN long range nodes was deployed in a rural area. It has been active since 2005 although the network topology has changed as it has been upgraded with new versions of the developed protocols. Fig. 12 shows the topology and traffic distribution of an early deployment during a 4 month period between 2005 and 2006 [170]. The network comprised 19 nodes and covered 2 km\(^2\) area.

The results highlight the unpredictability and dynamic nature of a WSN, even when nodes are stationary. The communication range can be very short due to environmental obstacles but also unexpectedly long due to signal reflections. Thus, the communication topology can be unpredictable as evident from Fig. 12. Link quality can change within few moments e.g. due to moving objects that block signal path, within few minutes or hours e.g. due to due to wet leaves and rain, or the quality can vary slowly based on the season of the year e.g. due to snow and lack of leaves on trees on winter. This necessitates the dynamic operation of network protocols.

![Fig. 12. Distribution of transmitted traffic on selected nodes in the long range outdoor deployment.](image-url)
4.4.2 Outdoor Environmental Monitoring in Suburban Area

An outdoor network with the TUTWSN long range platform was deployed in the suburban environment around Tampere University of Technology campus area. The network comprised 20 nodes that were deployed on 1 km long line topology. Thus, the goal was to experiment the effects of a multihop topology. The nodes were attached to trees as shown Fig. 13.

The deployment further highlighted the effect of environmental changes. In addition to the changing weather, vehicles cause network dynamics. For example, a parked truck can obstruct a link necessitating the discovery of alternate routes.

4.4.3 Indoor Deployment at TUT Campus

An indoor TUTWSN deployment has been used in Tampere University of Technology since 2008 for research and education purposes. Students utilize the network in an organized WSN course by designing sensor network applications that utilize the collected data. In total, the network comprises over 200 nodes in several campus area buildings and has a total network coverage of 23000 m². The nodes are attached to walls as shown in Fig. 14. Deployed nodes include temperature, humidity, illumination, Carbon dioxide (CO2), sound level, and Passive Infra-Red (PIR) sensors

Fig. 13. Node installation on trees in the outdoor deployment.
allowing e.g. air quality monitoring and space usage applications.

The large scale and long term deployment has highlighted the importance of energy-efficiency, scalability and autonomous operation of the network protocols, and the necessity of network diagnostics. For feasible network maintenance, the lifetime on batteries must be several years and network must be auto-configurable with multi-hop routing support. However, as node break ups, energy depletion, and wireless interference due to other networks are common in the large scale deployment, network diagnostics was identified as a necessity which further guided the development of diagnostics presented in this Thesis.
4. TUTWSN Platform and Deployments
5. QOS ANALYSIS FOR WSNS

The results of this Thesis are summarized in this and the following chapters. This chapter presents QoS metrics to measure network performance and presents a practical example of QoS trade-offs in TUTWSN protocols. The research related to this chapter is published in [P4].

5.1 QoS Metrics

Due to varying application requirements, QoS cannot be assessed as a single grade but needs to be described with a collection of several metrics. While the traditional reliability, latency, and throughput metrics apply to the WSNs, other metrics are required to comprehensively assess QoS. The term metric in this Thesis means a parameter that quantifies a certain aspect of network performance. From the protocol design point-of-view, some of the presented metrics may act as constraints instead. In this Thesis, a metric is understood to be a measurable goal for QoS, while a constraint is a limiting factor. For example, communication range is typically a constraint for a routing protocol. However, from the application and user point-of-view, the communication range is still a metric as it dictates, e.g., the number of nodes required to cover a certain area.

As the amount of potential WSN QoS metrics is large, this Thesis concentrates on the metrics that affect end-to-end traffic, and thus, the practical performance available for end users and sensor applications. The performance is considered at network and node levels, where the network performance is understood as an aggregate (e.g. average) of individual node performances. For quantitative QoS comparison, each QoS metric is assigned with a value and unit. The metrics are described in the following sections and summarized in Table 7.
Table 7. QoS metrics considered in this Thesis.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Unit</th>
<th>Metric</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>s</td>
<td>Throughput</td>
<td>bps</td>
</tr>
<tr>
<td>Reliability</td>
<td>%</td>
<td>Availability</td>
<td>%/s</td>
</tr>
<tr>
<td>Mobility</td>
<td>m/s</td>
<td>Lifetime</td>
<td>days</td>
</tr>
<tr>
<td>Communication range</td>
<td>m</td>
<td>Node count</td>
<td>pcs</td>
</tr>
<tr>
<td>Node density</td>
<td>1/m²</td>
<td>Security</td>
<td>(grade)</td>
</tr>
</tbody>
</table>

5.1.1 Latency

Latency denotes the elapsed time between the generation of a packet at a source node to its reception at the target node.

5.1.2 Throughput

The throughput metric expresses the amount of application payload transferred per time unit from a source to the target. In practice, the throughput is significantly less than the nominal transceiver data rate due to the protocol overhead and low duty cycling.

5.1.3 Reliability and Availability

The reliability metric denotes the probability that a packet is successfully delivered from a source to a target. Publication [P4] evaluates the effect of beacon losses, limited buffer space, and reasons for unreliability in WSNs.

Three distinct reasons for unreliability can be identified as shown in Fig. 15. First, a packet may be dropped due to link errors. The protocol design choices such as re-transmissions on MAC layer or the use of store-and-forward mechanisms on routing layer reduce the packet drops. Second, limited queuing space cause packet drops. As the typical data memory of a resource constrained WSN node is 2-8 kB [84], part of which is required by applications and protocol stack, the remaining buffer may overflow on a temporary traffic burst or when a next hop link is broken. This necessitates queuing disciplines and traffic differentiation to prevent the loss of high priority messages. Third, node failures due to hardware malfunction or depleted energy cause the loss of queued data. The recovery from these require end-to-end retransmissions or storing routed packets in persistent memory.
Due to the inherent redundancy of sensor data, the reliability metric is not always important in WSNs. Instead, it is important to ensure that the received sensor values are up-to-date. For this purpose, this Thesis defines the availability metric. The availability is defined as the probability that data is received from a node within certain time interval $I$ as

$$
\text{availability} = \frac{|\{1 \leq i \leq N-1 : a_i - a_{i-1} \leq I\}|}{N-1},
$$

(1)

where $a_i$ is the arrival time of the $i$th sample, $N$ is the number of received samples. Thus, a node is considered available when sensor values are received from it over the time of observation.

An example of the availability when a node generates traffic at constant 60 s intervals is shown in Fig. 16. From (1) it follows that the measured availability resembles the Cumulative Distribution Function (CDF) of the process that generates traffic at a source node. Therefore, when a packet is not lost and delay jitter is minimal, the average reception interval equals to the average traffic generation interval (60 s). Latency jitter spreads CDF around the average traffic generation interval, thus increasing the time to reach over 50% availabilities. Also, packet losses increase the time between receptions and consequently decrease the availability.

In practice, the availability evaluate the applicability of the network for a certain purpose. For example, in a WSN is targeted at intruder detection should not be unavailable for a long time or otherwise an alert can be received too late. Thus, an availability (e.g. 99.99%) must be associated with a time interval, e.g. one minute. In a measurement network the interval may be several minutes or even hours if the observed phenomena changes slowly.
5. QoS Analysis for WSNs

![Graph showing availability metric](image)

**Fig. 16.** Availability metric expressing the probability that an update is received from a node within certain time interval. Packet drops and network errors decrease the availability.

### 5.1.4 Network and Node Lifetime

The lifetime is considered as a QoS metric due to its importance for several WSN applications. Furthermore, many of the other QoS metrics have a trade-off between lifetime, thus preventing optimizing all metrics. For example, the typical energy saving mechanisms, such as low duty cycling, have a negative impact on throughput and latency. The lifetime of a node is defined as the elapsed time from its deployment to the depletion of its energy sources. The network lifetime is defined as the minimum lifetime of its nodes.

### 5.1.5 Node Density, Count, and Communication Range

Node density, node count, and communication range describe how a network can be deployed. Node density defines the maximum number of nodes that can operate within the communication range. Contention-free and beacon-enabled protocols typically necessitate non-overlapping data exchange times, thus having a design trade-off between the node density and the communication range.

Node count defines the maximum number of nodes in a network. While the number of nodes is ideally only limited by the network throughput, practical protocols may have design choices, e.g. in address assignment, that limit the node count.
5.1.6 Mobility

The mobility metric describes how fast a node can move in a network but still exchange data with other nodes. It is particularly important in tracking WSNs where a node may be attached to moving objects. Mobility can be improved by increasing communication range for longer link lifetimes and by increasing protocol reactivity that allows rapid neighbor discovery and communication links establishment.

An upper limit for mobility can be calculated as $R / (t_d + t_a + t_m)$, where $R$ is the communication range, $t_d$ is neighbor discovery time, $t_a$ is the negotiation time between a source and a target to establish communication link (e.g. association procedure), and $t_m$ is the time needed to transfer a message. In the case of low duty cycle protocols, the initial sleeping delay decreases mobility significantly. As an example, assuming beacon synchronized MAC with 2 s access cycle, the average sleeping delay is 1 s.

5.1.7 Security

Security means that unauthorized parties do not gain access or tamper with the sensed data [40]. As the sensor networks might carry sensitive sensor data or support actuation, security might be an important element of QoS. However, unlike the other metrics, security does not have a straightforward unit or value. For comparison purposes, the security should be graded based on supported security features.

As an example, Table 8 presents a simple method to grade network security. Each grade is an incremental improvement over lower grades to allow comparison between grades. The use of encryption ensures data confidentiality. However, freshness counters are required to prevent injection of recorded packets that could otherwise allow e.g. actuation. Data encryption can be either network wide with pre-shared key and algorithm, or negotiated per node. Network wide encryption is less secure because a revealed key compromises the whole network.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Encryption</th>
<th>Freshness counter</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>Network wide shared key</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Network wide shared key</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Node specific key</td>
<td>Yes</td>
</tr>
</tbody>
</table>
5.2 Usage of QoS Profiles

Together, a set metric values form a QoS profile. This can be illustrated with a radar chart as shown in Fig. 17. A larger surface area in the chart denotes better QoS. Thus, the chart allows comparing QoS e.g. between different networks, network configurations, and the network operation at different time instances. As some of the QoS metrics are only relevant for the whole network, such as node density, a QoS profile for a node contains a reduced set of metrics.

A prominent use case of the QoS profile chart is comparing the theoretical QoS (e.g. what a technology promises) against measured performance as shown in Fig. 18(a). In practice, the actual QoS is equal or lower than the theoretical due to both non-idealistic network environment causing unreliability and available capacity when network is not fully utilized. The second use case is the assessment of variations in network performance by taking the average, minimum, and maximum representatives of individual node metrics and presenting the cases in the same chart as shown in Fig. 18(b). A big difference between minimum and maximum values denotes potential performance problems. Third use case is to assess whether a technology can fulfill the application requirements, or to assess whether a deployed network operates as required. As an example, Fig. 18(c) presents a case where the measured performance indicates that all QoS requirements are not met.

![QoS profiles of (a) whole network and (b) a individual node.](image-url)
5.2. Usage of QoS Profiles

5.2.1 ZigBee Network Example

As a concrete example of the QoS profile usage, the performance of a ZigBee network comprising 32 nodes is analyzed in the following. The analysis assumes that the nodes form a symmetric tree topology, where each router (ZigBee coordinator) has \( n_c = 3 \) child routers and \( n_d = 5 \) non-routing (leaf) devices. Each node generates a 20 B measurement packet to the sink every 20 s, resulting into 296 bps total offered load.

QoS is analyzed by keeping the application traffic requirements fixed, while examining different Beacon Order (BO) (access cycle length) and Superframe Order (SO) (active period length) parameter values of IEEE 802.15.4. The parameters are listed in Table 9.

The end-to-end latency is evaluated for the leaf nodes that are farthest away (3 hops) from the sink by assuming that each hop causes on average an delay of 0.5 access cycles. The assumptions is fairly accurate as the access cycle length is significantly larger than the packet transmission times and the network is not saturated.

The throughput and reliability metrics are analyzed with the models from [85]. The

![Comparison methods for QoS](image)

**Fig. 18.** Comparison methods for QoS, a) theoretical vs. measured performance, b) performance differences between nodes, and c) application requirements vs. measured network performance.

<table>
<thead>
<tr>
<th>BO</th>
<th>SO</th>
<th>Active period (s)</th>
<th>Access cycle (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>0.246</td>
<td>0.989</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.123</td>
<td>1.97</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.246</td>
<td>3.93</td>
</tr>
</tbody>
</table>
used models consider collisions from hidden nodes and the contention-based channel access as the main sources for reliability (goodput) loss, and e.g., link errors, are not considered.

The mobility metric is calculated with the formula proposed in Section 5.1.6. As the IEEE 802.15.4 standard leaves several implementation aspects to the discretion of the equipment manufacturer, the following behavior is assumed. First, to ensure that a beacon from a neighbor is detected, each channel must be scanned (at least) for the duration of one access cycle. Next, a node must wait until the beginning of the next active period before an association can be requested, requiring an average wait time of \((\text{access cycle length} - \text{active period length})/2\). Assuming only one operating channel, discovery \((t_d)\) and association \((t_a)\) take approximately 1.5 access cycles. At least one data transmission is assumed before a node moves outside of the communication range. Communication range \(R\) is assumed to be 50 m. In the analysis, ZigBee tree-based routing is assumed. The use of mesh routing would require additional route messaging before data can be transmitted.

Energy consumption is evaluated with the models presented in [P6]. The settings of the High Rate (HR) platform presented in Table 2 of the publication were used, with the exception of data rate that was configured to 250 kbps to conform the IEEE 802.15.4 standard. The lifetime was calculated for the routers one hop from the sink, as these forward most traffic and has thus the highest energy consumption. 3000 mAh battery was assumed in the lifetime calculations. The full use of IEEE 802.15.4 data confidentiality, data authenticity, and replay protection features is assumed, and the security is graded as 3 based on Table 8. As the modeled reliability was over 99.5%, the availability is assumed be the same as the packet generation interval.

The results are presented in the Fig. 19 and Table 10. While all settings fulfill the application throughput requirement, they have trade-offs with latency, lifetime, and reliability. Thus, the parameters should be selected based on application demands.

### 5.3 QoS Metrics in TUTWSN

This section presents the configuration of TUTWSN as an example of QoS trade-offs in a WSN protocol. The parameters and their affect on QoS are listed in Table 11. The effects are listed with the assumption that the network is not saturated.
5.3. QoS Metrics in TUTWSN

Latency
Throughput
Reliability
Availability
Mobility
Lifetime
Comm. range
Security
SO4/BO6
SO3/BO7
SO4/BO8

Fig. 19. ZigBee QoS profile with selected IEEE 802.15.4 MAC settings.

Table 10. ZigBee QoS results with selected IEEE 802.15.4 MAC settings

<table>
<thead>
<tr>
<th></th>
<th>SO4/BO6</th>
<th>SO4/BO8</th>
<th>SO3/BO7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency (s)</td>
<td>2.2</td>
<td>8.6</td>
<td>4.3</td>
</tr>
<tr>
<td>Throughput (bps)</td>
<td>2487.8</td>
<td>622.0</td>
<td>309.8</td>
</tr>
<tr>
<td>Reliability (%)</td>
<td>99.6</td>
<td>99.6</td>
<td>99.5</td>
</tr>
<tr>
<td>Availability (s)</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
</tr>
<tr>
<td>Mobility (m/s)</td>
<td>37.0</td>
<td>8.7</td>
<td>17.3</td>
</tr>
<tr>
<td>Lifetime (days)</td>
<td>25</td>
<td>102</td>
<td>101</td>
</tr>
<tr>
<td>Comm. range (m)</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Security (grade)</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

5.3.1 Application layer

An application samples sensors with a certain measurement rate and transmits the samples to a sink. Increasing the measurement rate increases traffic load thus decreasing lifetime and available throughput. However, it also increases availability as data is received more often and reliability as possible packet losses are compensated by a new sensor sample.

Instead of sending measurements immediately, they can be aggregated by combining several values (e.g. averaging) or inserting several measurement values into a single packet. This decreases traffic overhead but increases latency and decreases availability as several measurements are cached before sending the aggregated value. Also, a


<table>
<thead>
<tr>
<th>Configuration / change</th>
<th>Comm. range</th>
<th>Lifetime</th>
<th>Mobility</th>
<th>Reliability</th>
<th>Availability</th>
<th>Latency</th>
<th>Throughput</th>
<th>Node density</th>
<th>Area coverage</th>
<th>Security</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Application layer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement rate / increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregation / increase</td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><strong>Routing layer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternative routes / increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Multipath routing / increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td><strong>MAC layer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access cycle length / increase</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>CAP length / increase</td>
<td>-</td>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>CFP length / increase</td>
<td>+</td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Network beacon rate / increase</td>
<td>-</td>
<td></td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Acknowledgments / use</td>
<td>-</td>
<td></td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Retransmissions / increase</td>
<td>-</td>
<td></td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Encryption / use</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><strong>Physical layer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency / increase</td>
<td>-</td>
<td>-</td>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Transmission power / increase</td>
<td>+</td>
<td>-/+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Data rate / increase</td>
<td>-</td>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

* + positive effect on QoS − negative effect on the QoS metric of the column*

packet loss causes the loss of several samples.

### 5.3.2 Routing layer

The use of alternative routes towards a sink requires maintaining synchronization with several neighbors which increases the energy usage. The benefit is that a replacement route is ready if a link breaks thus reducing the risk of buffer overflows. TUT-WSN also defines multipath routing where a packet is transmitted via each known route. This increasing reliability and mobility but requires more bandwidth and energy.

### 5.3.3 MAC layer

In TUTWSN, the superframe structure has the most significant effect on the MAC layer performance. Because a cluster beacon is sent every access cycle, a long access
cycle reduces energy usage. However, it also increases the forwarding latency since a member node must wait longer to send its data. Also, assuming that the active period length (CAP + CFP) is kept the same, duty cycle decreases and reduces the throughput. On the other hand, a low duty cycle allows fitting several non-overlapping clusters into the same channel thus increasing node density.

Increasing the CAP length increases idle listening on a cluster head but decreases collision probability, therefore increasing reliability. Also, as mobile nodes can communicate with a cluster only briefly before moving outside the communication range, long term reservations are not feasible and a long CAP increases mobility. A mobile node also requires a high network beacon rate for detecting new neighbor clusters rapidly.

The MAC layer recovers from failed transmissions with acknowledgments and retransmissions. However, as data frames are usually small and thus comparable to the acknowledgment frames, the use of acknowledgments essentially doubles the data transmission energy. Generally, not using acknowledgments increases throughput on reliable links but the throughput is unaffected in TUTWSN due to the slotted channel access. The retransmissions increase latency, because new frames have to be wait until an old frame is retransmitted.

In TUTWSN, security is implemented at the MAC layer by encrypting all data transmissions. This has a small effect on lifetime due to increased frame processing times and to the throughput as encryption adds a small communication overhead.

### 5.3.4 Physical layer

A transceiver has three important properties that affect the network performance: frequency, transmission power and data rate. These have a complex relations due to both physical properties and legislation restrictions e.g. necessitating lowering data rate to fit into the allocated frequency band. A detailed analysis on their relations is outside the scope of this Thesis. The effect of other physical layer components, e.g. MCU, is smaller and ignored in this example.

According to Friis communication equation [50], communication range decreases as frequency is increased. The communication range has a direct effect on mobility, area coverage, and node density. It can be further increased with high transmission power with the trade-off to the lifetime. It should be noted, however, that in low power radios the power consumption does not rise linearly with the output power. As an example, Table 12 presents measured communication range of the Nordic nRF905 [123] and
Table 12. Transmission power vs. range in TUTWSN platforms.

<table>
<thead>
<tr>
<th>Radio</th>
<th>Frequency (MHz)</th>
<th>Data Rate (kbps)</th>
<th>Output power (mW)</th>
<th>RX sensitivity (dBm)</th>
<th>Supply current (mA)</th>
<th>Comm. range (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nRF905</td>
<td>433</td>
<td>50</td>
<td>10.0</td>
<td>-100</td>
<td>30.0</td>
<td>500</td>
</tr>
<tr>
<td>nRF905</td>
<td>433</td>
<td>50</td>
<td>0.10</td>
<td>-100</td>
<td>9.0</td>
<td>105</td>
</tr>
<tr>
<td>nRF24L01</td>
<td>2400</td>
<td>1000</td>
<td>1.00</td>
<td>-85</td>
<td>11.3</td>
<td>180</td>
</tr>
<tr>
<td>nRF24L01</td>
<td>2400</td>
<td>1000</td>
<td>0.02</td>
<td>-85</td>
<td>7.0</td>
<td>10</td>
</tr>
</tbody>
</table>

nRF24L01 [122] transceivers in the TUTWSN platforms. With these transceivers, the highest output power has 4.8 and 18 times the range but only 3.3 and 1.6 times the supply current requirement, respectively. Thus, as fewer hops are required, the selection of higher transmission power can be feasible from the whole network point of view.

The data rate parameter has a trade-off between reliability and lifetime. Although a faster data rate typically slightly increases the transmission power per time unit, the overall energy requirement is lowered as more data can be sent at the same time interval. However, faster data rates often reduce reliability, as the sensitivity at the receiver decreases. A higher carrier frequency enables the use of wider communication band and therefore higher throughput.
6. PROTOCOL DESIGNS FOR QOS

This chapter summarizes the research results on MAC and routing protocol designs for QoS and presents a cross-layed design between the presented protocols. The details of the protocols are presented in [P1,P2,P5,P6].

6.1 QoS Schemes for WSN MACs

This research proposes two alternative QoS schemes for MACs. Both schemes support QoS classification and reservations. The first scheme, QoS support layer for WMNs [P5], is targeted at contention-based MAC protocols. The support layer can be realized without modifications to existing MAC protocols. The second scheme, dynamic capacity allocation [P2,P6], is targeted at MAC protocols with contention and contention-free channel access. It has more efficient channel usage and thus better energy-efficiency but requires tighter integration to the MAC layer.

6.1.1 QoS Support Layer for WMNs

The QoS support layer uses bandwidth management and admission control techniques that avoid saturating the communication channel. It relies on the fact that contention-based MACs support strict QoS requirements when offered traffic load is controlled and not near the maximum capacity [211].

The QoS support layer assumes a clustered topology, where the cluster head manages traffic within one hop radius. Cluster’s member nodes connect to the cluster head as depicted in Fig. 20.

A physical connection between two nodes consists of one or more logical links, each using distinct QoS definition that comprises operation mode, bandwidth limit, and priority. A link operate in one of the two modes: bandwidth reserved or differentiated. In the bandwidth reserved mode, a link is guaranteed with certain throughput that is initially requested from the controller. A source node is responsible for limiting
its average traffic to the agreed bandwidth limit. The remaining capacity is divided proportionally among the differentiated mode links based on their assigned priorities. In the differentiated mode, a node polls the controller for a permission to send. The controller schedules requests and grants permissions to send for a certain time period, this way preventing congestion. The link priority affects the duration and urgency of the granted transmission time.

The throughput with different number of source nodes is shown in Fig. 21. Each node transmits three flows to the cluster head: 64 kbit/s real-time, 200 kbit/s best-effort, and 400 kbit/s background flow. The results are obtained with NS2 simulations of IEEE 802.11 WLAN with the Distributed Coordination Function (DCF) MAC and Direct Sequence Spread Spectrum (DSSS) PHY models. Data rate, backoff, and CW parameters were the default values specified in IEEE 802.11b standard. The simulation details can be found in [P5]. Real-time flows reserved the 64 kbit/s bandwidth, whereas the best-effort flows are assigned with differentiated service priority 2 and the background flows are assigned with differentiated service priority 1. The results show that the bandwidth reserved mode guarantees the requested throughput even when the offered throughput exceeds the capacity. Remaining bandwidth is divided between the differentiated flows.

The key innovation in the QoS support layer is the co-existence of different types logical links, efficient bandwidth usage, and low overhead: control messaging and polling is used only when the communication channel usage is near its saturation point. Also, unlike the fixed reservation schemes, unused reservations do not waste capacity as the excess bandwidth is assigned for the differentiated flows. The logical link based approach also enables end-to-end QoS flows. For example, links $b$, $c$, and $d$ in Fig. 20 might use similar QoS settings thus defining an end-to-end flow. However, the construction of such flows is performed with a higher layer protocol, such as Resource Reservation Protocol (RSVP) [19], and is outside the scope of this Thesis.
6.1. QoS Schemes for WSN MACs

6.1.2 Dynamic Capacity Allocation

The dynamic capacity allocation scheme assumes a beacon-enabled MAC that supports both contention-based and contention-free channel access. These assumptions are compatible with many existing WSN MACs such as IEEE 802.15.4. Unlike the related proposals that utilize contention-free channel access via static reservations, this scheme assigns contention-free slots dynamically based on traffic requirements. The scheme has three distinct benefits. First, the contention-free period is used only to manage reservations, therefore allowing to minimize its length and thus reducing idle listening. Second, contention-free period can also serve traffic bursts. Third, the scheme aims to minimize unused reservations with dynamic slot assignment.

The superframe used in the scheme is presented in Fig. 22. First, a cluster head transmits a beacon that describes the structure of the superframe. This is followed by short contention period that can be used to request reservations or send data if a contention-free slot is not granted. The following contention-free slots are assigned by the cluster head based on traffic requests.

Three types of contention-free services are defined:

**Guaranteed**: Members explicitly request certain amount of reservations from a cluster head. The service ensures certain minimum throughput.

**Adapted**: A cluster head records the average traffic usage of its member nodes and automatically assigns slot reservations to match the traffic. This way, the
the need for reservation signaling is reduced which in turn reduces the use of contention access period.

**On-demand**: A member node may request for an additional reservation in any transmission (contention based, adapted, or another on-demand) by setting a slot demand flag on the header of transmitted frame. Cluster head indicates the granted slot (or none if the superframe was full) in its acknowledgment.

The services operate seamlessly together. The guaranteed and adapted services are best suited for CBR traffic, while the on-demand service handles traffic bursts.

The amount of guaranteed and adapted slots is defined as slots per time unit e.g. slots per minute. Therefore, a member node might not receive the same amount of reservations each access cycle. For example, assuming 10 guaranteed slots per minute, no adapted slots, and 2 s access cycle, a member node receives guaranteed slot every third access cycle. This provides a trade-off between energy and forwarding latency to a node: it can either transmit data immediately with the contention based channel access or wait few access cycles until a contention-free slot is granted. By sending immediately the node risks collision and might not have anything to send when the cluster head next time grants the reserved slot.

The average delay caused by postponing the frame transmissions until a next contention-free slot is granted is

\[
delay = \min \left( \frac{R_A}{r}, T_A \right) \cdot t_{ac},
\]

where \(T_A\) is a configurable wait time for the contention-free slot, \(R_A\) is reservation period length, \(r\) is the total number of granted reservations per period, and \(t_{ac}\) is the access cycle length

Figure 23 shows simulated one hop latency with on-demand, guaranteed, and adapted capacity services, when the maximum reserved slot wait time \((T_A)\) is 4 s (two access cycles). The simulation parameters are described in detail in [P6]. In these results, the guaranteed and adapted services also supported on-demand allocations.
The on-demand service has the highest delay because it always waits $T_A$ access cycles for a contention-free slot. The delay decreases slightly on higher traffic loads (smaller data generation interval) as several packets are buffered and can be transmitted after the initial wait time. The adapted service has low latency on high traffic loads, because a slot is granted on every access cycle. On low traffic loads, the probability that a node is granted with a reservation is small, thus increasing the average waiting time. The guaranteed traffic service performs similarly regardless of the load because the fixed reservations also guarantee a certain upper limit for latency.

The power consumption with the different services is presented in Fig. 24. On a subnode, the guaranteed service consumes 4% more power than the other services when the traffic load is low. However, as a subnode forwards only its own traffic, the differences between the schemes are otherwise negligible.

The guaranteed service has the highest power consumption on low traffic loads also on a cluster head. Although the bandwidth reservations were adjusted based on the known traffic load, the probability that a node has data to send when a reservation is granted decreases as the data generation interval increases. Thus, the reservations are unused which causes unnecessary listening.

The on-demand service has the lowest power consumption because it completely avoids unnecessary reservations. However, its CAP usage was high, 42% with 1 s data generation interval, whereas other methods had only 2% load. Thus, the length of contention-free period could be reduced from the fixed two slots ($N_A = 2$) when using the other methods. Using only one contention-based slot ($N_A = 1$) with the dynamic service decreases its power consumption by 24%, while the same amount of slots would congest on-demand service. Decreasing the number of contention slots
with the dynamic service changed the per-hop latency and subnode power consumption less than 1%.

The results highlight the main benefit of the allocation scheme: the length of the contention-free period can be reduced thus significantly decreasing the energy usage of a cluster head. As the cluster head consumes most energy, this increases the lifetime of a network. Also, the results indicate that the scheme offers trade-offs between latency, energy-efficiency, and capacity.

### 6.2 QoS Routing Cost Algorithm

This section presents an algorithm for QoS route selection and cost calculation published in [P1]. As several practical routing protocols (e.g. ZigBee) calculate route cost, the algorithm is applicable to several existing routing protocols. Unlike other proposed cost algorithms, the routing proposed in this Thesis uses several QoS metrics to calculate the routing cost. These metrics are reliability, throughput, delay, and energy. Other QoS metrics, such as security or mobility, are less influenced by the choice of routing path and thus not included. Fig. 25 summarizes the utilized QoS metrics and the statistics that are used to calculate the cost.

#### 6.2.1 Minimum Cost Routing

Route selection assumes that a) cost is increased on each hop and b) the route that has the lowest cost is selected. If two routes have similar QoS, e.g. in respect of
6.2. QoS Routing Cost Algorithm

Fig. 25. QoS metrics and statistics used to derive cost in TUTWSN routing.

reliability, the route that is shorter has a lower cost and is thus selected. For comparison, the route cost can be calculated by multiplying link reliabilities along a routing path but this increases routing overhead as either end-to-end route construction or an additional hop counter is required to prevent loops.

The cost $C_i$ of node $i$ when routing via node $j$ is expressed as

$$C^i = C^j + C^{i\rightarrow j},$$

where $C^j$ is the cost advertised by node $j$ and $C^{i\rightarrow j}$ is an additive cost. The additive cost denotes both link cost used to avoid bad links and node based cost at $i$th node used to discourage other nodes from routing via a node, e.g. when a node has low energy. To ensure that the cost is increasing, $C^{i\rightarrow j}$ must have a positive non-zero value.

6.2.2 Cost Algorithm

To construct the additive cost between nodes ($C^{i\rightarrow j}$), QoS routing first evaluates link quality with cost functions at time $t$:

- energy $e_{ij}(t)$ represents the energy needed to forward a packet from node $i$ to node $j$,
- reliability $r_{ij}(t)$ that denotes the forwarding success rate between nodes $i$ and $j$,
- delay $d_{ij}(t)$ denoting the average time until a packet has been successfully forwarded to node $j$,
- residual energy $E_i(t)$ is the energy at node $i$, and
traffic load \( L_i(t) \) at node \( i \).

For brevity of the presentation, time instance \( (t) \) is not included when discussing the cost functions in the remainder of this Chapter. As functions relate to different QoS metrics that cannot be compared against each other as such, the cost function return unitless values that are scaled within the same range \([0, 1]\). A higher value means a negative effect on QoS. For example, \( r_{ij} = 1 \) means that all packets on link are lost, \( e_{ij} = 1 \) means that a high transmission power or several retransmissions are required to deliver packet to the next hop, \( E_i = 1 \) denotes that node’s energy source is depleted, and \( L_i = 1 \) means that the available bandwidth or the packet buffers of the node \( i \) has been exceeded.

Next, the values of cost functions are multiplied against cost weights as shown in Fig. 26. Thus, the cost weights describe which QoS metrics are considered important. Finally, the weighted values are combined to a single cost value that acts as the additive cost:

\[
C^{i\rightarrow j} = w_e \cdot (0.5 \cdot e_{ij} + 0.5 \cdot (1 - E_i)) + w_r \cdot r_{ij} + w_d \cdot d_{ij} + w_t \cdot L_i, \tag{4}
\]

where \( w_e, w_r, w_d, \) and \( w_t \) are cost weights for energy, reliability, delay, and throughput. Using two cost functions that relate to energy \((e_{ij} \text{ and } E_i)\) allows selecting between the least energy consuming route and balancing the energy consumption within the network [186]. Selecting only minimum energy route may cause early depletion of heavily loaded nodes, whereas considering only the residual energy can consume more energy globally, thus shortening the total network lifetime [24]. The optimal selection between the two energy calculation techniques depends on application requirements and network topology. As a compromise, they are weighted equally in this calculation.

---

**Fig. 26.** QoS metrics and statistics used to derive cost in TUTWSN routing.
The cost weights are defined as

\[ 1 = w_e + w_r + w_d + w_t. \]  

(5)

Thus, when one weight is increased, other weights must be decreased. This way, there is a clear trade-off between different metrics.

6.2.3 Cost Functions for TDMA-based MAC

This section defines cost functions for TDMA-based low duty cycle MACs such as TUTWSN MAC. Defining the functions for other MAC protocols is outside the scope of this Thesis.

Load \( L_i \) is calculated from the amount of utilized slots versus the maximum number of slots. Residual energy \( E_i \) is derived from node voltage measurements, where \( E_i = 0 \) equals to the typical maximum battery voltage and \( E_i = 1 \) equals to the shutdown voltage.

Other components are more complex as they need to consider packet error rate, denoted as \( p_{ij} \), in addition to the component specific parameters. Packet error rate causes retransmissions that have an effect to the energy and delay. As practical MACs have an upper limit \( u_{\text{max}} \) for the attempts, the number of transmissions on link between nodes \( i \) and \( j \) is formulated as

\[ u_{ij} = \min \left( \sum_{k=0}^{\infty} k \cdot (1 - p_{ij})^{k-1}, u_{\text{max}} \right) = \min \left( \frac{1}{(p_{ij})^2}, u_{\text{max}} \right). \]

(6)

The result of the sum is derived from geometric series.

Two sources for unreliability were identified in [P4], transmission errors and buffer overflows. As the link and queue drop probabilities are independent, reliability component is derived from packet error rate \( p_{ij} \) and queue drop probability \( q_i \) at node \( i \) as

\[ r_{ij} = (1 - p_{ij}^{u_{ij}}) \cdot (1 - q_i). \]

(7)

It should be noted that these models assume uncorrelated packet losses. Generally, this is not the case as it has been observed that a wireless channel may enter a state where errors occur in bursts for a small interval [11]. However, as the error state has been observed to last less than few tens of milliseconds [115], the time between consecutive packet transmission in low duty cycle networks is much longer than the
channel coherence time. Therefore, the model is well suited for low energy WSN techniques and averages temporary performance variances.

The forwarding energy function $e_{ij}$ comprises the energy required to transmit a frame ($E_{tx}$) and receive an acknowledgment ($E_{rx}$). These can vary depending on the link quality and configured transmission and reception power levels. Thus, the energy is expressed as

$$e_{ij} = \frac{u_{ij} \cdot (E_{tx} + E_{rx})}{u_{max} \cdot E_{trx,max}}.$$  (8)

The fraction in the formula is used to normalize $e_{ij}$ to range $[0, 1]$. $E_{trx,max}$ is the maximum energy consumption with the highest transmission power and receiver sensitivity. Without power level adjustments, $E_{trx,max}$ equals to the $E_{tx} + E_{rx}$.

Delay function is derived by assuming low duty cycle operation where the propagation and channel access delays are small compared to the sleeping delay ($D_s$) and can therefore be ignored. The modeled delay consists of an initial sleeping delay $D_s$ until a packet transmission can be attempted first time. In addition, retransmission might postpone transmission until next access cycle as expressed in

$$d_{ij} = \max(D_s + \frac{u_{ij} \cdot D_{ac}}{v} \cdot D_q) \cdot \frac{1}{D_{max}},$$  (9)

where $v$ is an estimated number of transmissions per access cycle, estimated from the typical number of reserved slots per access cycle, and $D_{ac}$ is the access cycle length. The fraction in formula normalizes the value of the function to an estimated maximum delay $D_{max}$ value.

### 6.2.4 Simulation Results

The QoS routing cost algorithm was simulated with TUTWSN MAC on NS2 (version 2.31). In these tests, a sink broadcast a route advertisement once per 60 s. A node recomputed its cost when receiving the advertisement and forwarded the advertisement with an updated cost. In the simulations, an active period consisted of 16 contention-free slots, resulting 1.0 kb/s maximum throughput with the used 32 B data frame size and 4 s access cycle. MAC was configured with guaranteed service granting 2 slots per access cycle but a node could use on-demand service to request more bandwidth.

For realistic results, the transceiver power consumption was modeled after Chipcon CC2420 transceiver [180], which is a commonly used ZigBee compliant WSN trans-
6.2. QoS Routing Cost Algorithm

The simulations used transmission power level adjustment, because it is a recommended practice in wireless networks as it both decreases energy usage and reduces interference. The minimum power level that resulted into the average transmission success probability of 99% was used. As waking up a transceiver from the sleep typically takes some time, the transceiver was set to idle mode during an active period and to sleep mode during an idle period. The power consumption in different modes is summarized in Table 13.

Shadowing propagation model was used to model an obstructed office environment by setting shadowing deviation parameter to 4 dB and path loss exponent parameter to 4.95. In the shadowing model, the link error rate depend on the distance between nodes, which is the realistic behavior. The used parameters give a reliable connection with 0 dBm transmission power when distance is less than 28 m. The reception probabilities in 30 m, 40 m, 50 m, and 60 m distances are 99%, 70%, 14%, and 0.1%, respectively.

The simulation topology consisted of 50 nodes placed on a grid. The distance between adjacent nodes was 28 m. Thus, the nodes in the center of the grid had 4 neighbors within 28 m distance and 4 additional neighbors within 40 m distance. Three of the nodes were configured as sinks, while 18 of the nodes transmitted data to the sinks (6 source nodes per sink). The other nodes acted as routers. To gain confidence on the results, the results were averaged over 10 repetitions with different source nodes. As the hop count affects latency, the source nodes were selected so that their average distance to the sinks was 70 m. This way, the average hop count was the same due to the grid topology.

To show the difference between cost components, each cost weight ($w_e$, $w_r$, $w_d$, and $w_t$) was maximized separately. These were compared against max-min energy routing [98] and the cost calculation used in ZigBee [215]. Max-min energy routing

<table>
<thead>
<tr>
<th>Mode</th>
<th>Power consumption (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmit (-5 dBm)</td>
<td>42.0</td>
</tr>
<tr>
<td>Transmit (0 dBm)</td>
<td>52.2</td>
</tr>
<tr>
<td>Receive</td>
<td>56.4</td>
</tr>
<tr>
<td>Idle</td>
<td>1.28</td>
</tr>
<tr>
<td>Sleep</td>
<td>0.192</td>
</tr>
</tbody>
</table>
defines the routing cost for a route along nodes \( n_1, n_2, \ldots, n_{k-1} \) as

\[
C_R = \max_{i=1}^{k-1} \frac{1}{E_i},
\]

where \( E_i \) is the residual energy at node \( i \). In ZigBee, an implementation may report 7 as cost value or derive the cost from the probability of packet reception on a link \( (p_l) \) as

\[
C\{l\} = \begin{cases} 
7 & \text{min}\{7, \text{round}(1, p_l^4)\} \\
\end{cases}
\]

In these simulations, the reliability was used.

Figure 27 shows the average end-to-end reliability, end-to-end latency, and network lifetime. End-to-end reliability and latency were calculated as an average of individual values of the data transmitting nodes. End-to-end reliability does not reach 100% as each link has non-zero packet error rate. Instead, a packet is dropped after its transmissions attempt limit is exceeded or when a link breaks due to too many consecutively missed synchronization beacons. The simulations used 4 attempts which is the typical value in many practical MACs, e.g. IEEE 802.15.4 [70]. The network lifetime was defined as the time until the first node depletes its energy.

The best reliability, latency, and lifetime is achieved by maximizing the respective weight \( (w_r, w_d, \text{or } w_e) \). When the traffic load is high, maximizing the throughput weight \( (w_t) \) gives the smallest latencies and the best end-to-end reliability as the least loaded routes are selected.

The proposed reliability cost function \( (w_r) \) has 99.4% end-to-end reliability on low traffic load. In comparison, the ZigBee cost function has 98.8% end-to-end reliability. The reliability of the proposed cost function is higher because it evaluates the actual link reliability after retransmissions instead of using an arbitrary exponent. In addition, the proposed cost function also considers queue drop probability, thus having over 10% higher end-to-end reliability when the traffic load is high.

Compared to the max-min energy routing, the proposed energy cost function \( (w_e) \) gives 1% . . . 6% longer network lifetime depending on traffic load. While theoretically the max-min energy routing should maximize the time until first node dies [98], in a practical network with packet errors the max-min energy routing selects unreliable links and wastes energy due to retransmissions.
6.3 Cross-layer Design

The cross-layer design in TUTWSN integrates the presented dynamic capacity allocation algorithm and QoS routing protocol. In practice, the cross-layer design is realized with shared traffic classes and packet queue as shown in Fig. 28.

TUTWSN defines a configurable traffic class that defines the QoS. A sensor application selects the most suitable traffic class for its data and identifies the traffic class in the transmitted packet. The overhead is small as a class can be expressed with only few bits. For example, in the TUTWSN implementation, two traffic classes were used which required one bit overhead.

The traffic class comprises routing weights, traffic priority, and aging time. Routing weights are used in the route cost calculations. In addition, the maximum number of access cycles to wait for a CFP slot at the MAC layer (used by the dynamic capacity allocation algorithm) is derived from the routing weight for delay ($w_d$) as

$$T_A = (1 - w_d) \cdot T_{A,\text{max}}.$$  \hfill (12)
As a result, transmission is attempted immediately for traffic that is configured to prioritize latency \((w_d = 1)\), while other traffic may wait up to \(T_{A,max}\) access cycles. The traffic priority has two purposes. First, high priority packets are enqueued and transmitted first. Second, lower priority packets may be replaced when memory is full. Thus, the priority setting promotes low latencies and improves reliability for high priority packets when a network or node is congested. As such, high priority settings is intended for emergency and alert packets, whereas a lower priority is used for normal traffic. Aging time defines a maximum queuing time per hop until a packet is discarded. The parameter reflects the redundant nature of sensor data: if a route is unavailable until a new measurement is received, it might be better to discard the old measurement and deliver only the new one to save energy.

TUTWSN routing uses an interest based [72, 118] approach where a sink injects one or more data requests (interests) to the network [P1]. Nodes that can fulfill a request generate data and send it back to the sink. An interest defines

- target application or sensor that produces the desired information (e.g. temperature or diagnostics),
- area identifier or hop count from a sink to limit interest to certain part of the network,
- traffic class that should be used to send data,
- data generation interval, and
- sensor value range to generate data only when certain thresholds are exceeded.

As an example, normal temperature readings might use long data generation interval and normal traffic class to maximize energy-efficiency, while temperature readings exceeding fire alarm threshold would use much shorter data generation interval and high priority traffic class to minimize latency and maximize reliability.

In the implementation, traffic classes are programmed to selected default values dur-
Fig. 29. Traffic class and interest configuration via user interface.

The protocol stack was implemented in Microchip PIC18LF8722 MCU comprising only 128 kB program memory and 4 kB data memory. TUTWSN MAC protocol including the dynamic allocation algorithm consumed 37 kB program and 640 B data memory, while the routing protocol required 16 kB program and 170 B data memory. Two traffic classes were supported, necessitating only 1 bit overhead on data packets to denote the class. Route advertisements comprising sink address (3 B), route sequence numbers (route freshness counter, 1 B) and costs (1 B) for both classes caused the total of 7 B overhead. As the advertisements were piggybacked to cluster beacons,
they did not consume any additional overhead, Thus, the QoS based protocol designs presented in this Thesis can be implemented on very resource constrained hardware.
7. NETWORK DIAGNOSTICS

This chapter summarizes the research results on network diagnostics that can be used analyze realized QoS, and to detect and identify the performance issues.

7.1 Diagnostics Architecture

The diagnostics architecture comprises embedded self-diagnostics on sensor nodes, diagnostics data collection on a gateway, and data storage and analysis as shown in Fig. 30. An analysis tool refines and visualizes the collected information, while an alert service can notify e.g. via an email when battery depletes or a node disconnects [143]. The data is stored and analyzed in computer network due to the resource constraints and to allow maintaining longer history records. Still, the analysis could be located in embedded sensor nodes, e.g. to allow network self-configuration based on the diagnostics.

Unlike the related work [87, 106, 198] that concentrates on remote debugging of software failures and filtering failed sensor values, this work allows determining the cause for the performance problems. The collected information allows suggestions to add new nodes or move existing nodes to solve performance bottlenecks or unreliable links.

Fig. 30. Remote diagnostics collection architecture comprising embedded self-diagnostics on a sensor node, diagnostics data collection and storage, and an analysis tool.
7.2 Embedded Self-diagnostics

The embedded self-diagnostics aggregate statistics from different protocol layers. As these statistics are typically necessary to the node’s operation, the self-diagnostics incurs only small program and data memory overhead. The diagnostics data is passed to the gateway on application layer using the underlying protocol stack. Thus, the diagnostics does not require changing the communication protocols.

As all of the self-diagnostics information may not be needed at the same time, the diagnostics data is divided into several independent categories. Each category is associated with a certain collection period that determines how often the diagnostics data is sent to the gateway. Only the categories of interest are collected. In addition, different nodes may be instructed to collect a different set of diagnostics. Thus, the overhead, energy-usage, and the impact to other traffic can be minimized with selective diagnostics collection, making the approach feasible for the resource constrained WSNs. In practice, each category is transmitted to the gateway in a separate packet. The categories and collected statistics are summarized in Table 14.

7.2.1 Node information

Node information includes generic performance statistics and allows detecting performance problems that manifest as increased queue usage, node reboots, route changes, or network scans. Thus, the node diagnostics can be always active and used to switch on other, more extensive diagnostics when a symptom for a misbehavior is detected.

In addition to the performance diagnostics, node information includes remaining energy estimation used to determine when to replace the batteries. For practical reasons, the prototype implemented this with battery voltage. While the voltage indicates when the battery is about to deplete, it is inaccurate as an lifetime estimator due to non-linear relation between voltage and remaining energy. Thus, in many cases it would be preferable if a node could estimate its lifetime in percentage value or at real-time value.

7.2.2 Network and node events

Events assign a reason for specific outcomes, e.g. a network scan event occurred because a next hop link was lost. This information is crucial for detecting and analyzing problems that cannot be expressed as simple counters.
7.2. Embedded Self-diagnostics

Table 14. Embedded self-diagnostics information grouped by category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Node</strong></td>
<td>Voltage</td>
<td>Latest voltage measurement</td>
</tr>
<tr>
<td><strong>Information</strong></td>
<td>Queue statistics</td>
<td>Average and maximum queue usage and delays</td>
</tr>
<tr>
<td></td>
<td>Role</td>
<td>Cluster head or member node</td>
</tr>
<tr>
<td></td>
<td>Boots</td>
<td>Boot counter</td>
</tr>
<tr>
<td></td>
<td>Network scans</td>
<td>Network scan counter</td>
</tr>
<tr>
<td></td>
<td>Route changes</td>
<td>Cumulative number of route changes</td>
</tr>
<tr>
<td><strong>Network and node events</strong></td>
<td>Event</td>
<td>The descriptor of an occurred event</td>
</tr>
<tr>
<td></td>
<td>Reason</td>
<td>A reason for the event</td>
</tr>
<tr>
<td><strong>Network Topology</strong></td>
<td>Neighbor</td>
<td>Neighbor identifier (e.g. unique address)</td>
</tr>
<tr>
<td></td>
<td>Link quality</td>
<td>Link quality indication</td>
</tr>
<tr>
<td></td>
<td>Channel</td>
<td>Frequency that the neighbor operates on</td>
</tr>
<tr>
<td></td>
<td>Sleep schedule</td>
<td>Duty cycle timing relative to the sender</td>
</tr>
<tr>
<td><strong>Cluster traffic</strong></td>
<td>Channel usage</td>
<td>Average and maximum channel usage</td>
</tr>
<tr>
<td></td>
<td>RX/TX counters</td>
<td>The number of attempted and failed operations</td>
</tr>
<tr>
<td><strong>Link traffic</strong></td>
<td>Neighbor</td>
<td>Neighbor identifier</td>
</tr>
<tr>
<td></td>
<td>RX/TX counters</td>
<td>The number of attempted and failed operations</td>
</tr>
<tr>
<td><strong>Activity</strong></td>
<td>MCU activity</td>
<td>Time spent in active and idle states</td>
</tr>
<tr>
<td></td>
<td>Radio states</td>
<td>Time spent in RX, TX, idle, and sleep modes</td>
</tr>
<tr>
<td><strong>Route</strong></td>
<td>Path</td>
<td>List of forwarding nodes</td>
</tr>
<tr>
<td><strong>Routing latency</strong></td>
<td>Latency</td>
<td>End-to-end latency</td>
</tr>
<tr>
<td></td>
<td>Energy</td>
<td>Consumed energy to forward a packet</td>
</tr>
<tr>
<td></td>
<td>Hop count</td>
<td>Number of unique hops to the sink</td>
</tr>
<tr>
<td><strong>Software errors</strong></td>
<td>Boot reason</td>
<td>Last boot reason: assertion, low voltage, ...</td>
</tr>
<tr>
<td></td>
<td>Call stack</td>
<td>List of function addresses</td>
</tr>
</tbody>
</table>

7.2.3 MCU and transceiver activity

Activity diagnostic expresses the fraction of time spent in MCU active ($t_{mcu}$), radio reception ($t_{rx}$), and radio transmission ($t_{tx}$) states. It allows detecting unusually high transceiver or controller activity that might indicate other problems. In addition, the activity diagnostic allows estimating the average power consumption of a sensor node when static power consumptions of different operation modes is known. This approach is similar to the [41] but extended here to allow remote diagnostics. The
average power consumption $P$ is calculated as

$$ P = t_{mcu} \cdot P_{mcu} + (1 - t_{mcu}) \cdot P_{sleep}$$
$$ + t_{rx} \cdot P_{rx} + t_{tx} \cdot P_{tx} + (1 - t_{rx} - t_{tx}) \cdot P_{off}, $$  \hspace{1cm} (13)

where $P_{mcu}$, $P_{sleep}$, $P_{rx}$, $P_{tx}$, and $P_{off}$ are static power consumptions of MCU active, MCU sleep, radio reception, radio transmission, and radio off states, respectively. These are platform specific constants and can be stored e.g. in the diagnostics database, thus reducing overhead.

### 7.2.4 Route and routing latency

The route diagnostic describes end-to-end data forwarding. It contains the routing path as a list of node addresses. Each forwarding node updates the list. The information can be used to detect unusually long routes and allow understanding how the routes change over time. Routing latency describes end-to-end latency. Each forwarding node $n$ updates routing latency $t_n$ as

$$ t_n = t_{n-1} + (t_1 - t_0) + T_{toa}, $$  \hspace{1cm} (14)

where $t_{n-1}$ is the latency in a packet that is received from the previous hop $n-1$, $t_1$ is the forwarding time, $t_0$ is the reception time, and $T_{toa}$ is time-on-air that is estimated from the transceiver data rate and packet length. The latency information is a part of the route diagnostic but could also be piggybacked to data packets for continuous latency monitoring.

### 7.2.5 Cluster and link traffic

The hop-by-hop traffic is described with attempt and success counters of receptions and transmissions, which allows calculating link reliabilities and estimating the used bandwidth. The cluster traffic diagnostic describes the aggregate traffic flowing in and out from a node. The link traffic diagnostic is more descriptive as it maintains separate counters for each neighbor but has a higher overhead as the number of links is typically higher than the number of clusters in the network. Depending on the required level of detail, either diagnostic may be switched on.
7.2. Embedded Self-diagnostics

7.2.6 Network topology

Network topology describes the structure of the network and is essential when trying to determine how problems in one node can affect the rest of the network.

The neighbor information includes link quality (e.g. Received Signal Strength Indicator (RSSI)), channel, and sleep schedules. The link quality information approximates the relative distance between nodes, thus allowing more accurate comprehension on the topology. In addition, the information allows detecting when a node has only low quality links, which requires a user interaction to add new nodes in the vicinity to ensure reliable data forwarding.

The sleep schedules relate to the low duty cycle operation. Assuming that each node receives data on their own active period and forwards the data on the active period of a next hop neighbor, duty cycling incurs a significant forwarding delay. Considering these delays, it is possible to calculate the optimal routing delay between a node and a gateway. This allows detecting performance problems, when the optimal delay is compared against the actual diagnosed delay.

7.2.7 Software errors

Due to the resource constraints and tight coupling between software and hardware, embedded WSN devices are typically programmed with C or assembly languages that lack the advanced exception handling and memory overwrite protection of higher level languages. As a result, the embedded programming is error prone and some errors surface only in the actual deployments as the environment or network composition differs from the testing phase.

The proposed software diagnostics indicates the reason and the place in code where the problem occurred. As the information need to be transmitted only when a problem occurs, bandwidth is not typically required and the approach can be used in actual deployments to catch errors not found during testing. Two types of programming errors are covered. First, the diagnostics provides information on serious errors that prevent the execution of the embedded software, such as memory corruption, stack overflows, hardware failure, and other unexpected or unhandled events. This is realized by placing assertion statements to the code. Second, the self-diagnostics allows detecting infinite code loops with a software watchdog timer. Unlike the typical approach of using hardware watchdog timers to reboot the device, this approach allows identifying the problematic code segment.
If the boolean condition assigned to a assertion is false or software watchdog timer triggers, the self-diagnostics reboots the device to ensure a clean state as presented in Fig. 31. The diagnostics require a persistent memory, such as EEPROM, to maintain information while the node boots. The persistent memory holds an incremental boot counter, last boot reason, and the call stack of the executed program. This information is transmitted to a gateway after a boot.

### 7.3 Diagnosed QoS Metrics

The relation between the QoS metrics and collected diagnostics is summarized in Table 15. Security related diagnostics are not collected, as security features that affect the confidentiality, integrity, and authenticity of data, e.g., encryption, are often determined at deployment time and their change typically requires user intervention. However, it should be noted that jamming resistance against Denial-of-Service (DoS) attacks is another aspect of security [28] but this is reflected in availability. The node density can be calculated without diagnostics with known node locations. The diagnostics information allows evaluating QoS at two levels: network wide (end-to-end) or local (link or node). Evaluating network wide QoS gives an average network performance, while examining the local performance identifies performance bottlenecks.

### 7.4 Performance Analysis Tool

The performance analysis tool presents visually the collected self-diagnostics. The received node operation, link traffic (referred to as data forwarding in the screen

**Fig. 31. Triggering software diagnostics on assert or unintentional endless loop.**
Table 15. Relation between QoS metrics and collected diagnostics.

<table>
<thead>
<tr>
<th>QoS metric</th>
<th>Network</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>Worst node availability</td>
<td>Packet interarrival times at sink</td>
</tr>
<tr>
<td>Reliability</td>
<td>Missing sequence numbers</td>
<td>Transmission success counters</td>
</tr>
<tr>
<td>Latency</td>
<td>Routing latency</td>
<td>Queuing delay</td>
</tr>
<tr>
<td>Throughput</td>
<td>Received traffic at sink</td>
<td>Cluster/node traffic counters</td>
</tr>
<tr>
<td>Lifetime</td>
<td>Worst node lifetime</td>
<td>Long term: voltage drop</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Short term: MCU and radio activity</td>
</tr>
<tr>
<td>Mobility</td>
<td>Sum of route changes</td>
<td>Amount of route changes</td>
</tr>
<tr>
<td>Comm. range</td>
<td>Average node ranges</td>
<td>Estimate from neighbor info and known node locations</td>
</tr>
<tr>
<td>Node count</td>
<td>Active nodes</td>
<td>-</td>
</tr>
</tbody>
</table>

capture), cluster traffic diagnostics are directly visualized as charts and tables, but the tool also combines and refines the existing data for more descriptive information. Figure 32 shows the chart selection interface, which is preceded by network, examined node, and analyzed timeframe selection. The actual main interface of the diagnostics tool comprises a tabbed view of different charts.

The tool comprises three levels of analysis: availability analysis, advanced views
based on the self-diagnostics, and the presentation of sensor values. The availability analysis does not require self-diagnostics but can be evaluated from any received data. However, self-diagnostics is required for evaluating the performance. Sensor values allow examining the effect of environmental conditions, e.g. humidity affecting communication range, to the network operation.

7.5 QoS Analysis on an Outdoor Network

The developed diagnostics were used to analyze QoS in a network comprising 14 long range 433 MHz TUTWSN nodes. The nodes were deployed on 1.2 km long line topology to highlight the effects of a multihop topology.

Received packets chart shown in Fig. 33 is the basic method to examine network operation. The data point in the chart denotes a packet reception. Thus, a network problem manifests as a long period between packet receptions. In the figure, packet delivery to the sink (node identifier 233) was interrupted at 9:17. Based on event diagnostics, the reason was a broken link between the sink and the node that forwarded most of the network data.

The availability in the test network is presented in Fig. 34. The diagnostics tool vi-

![Fig. 33. WSN diagnostics UI showing received packets per node. The hole at 9:17 indicates network problems.](image-url)
visualizes availability with a summary table and chart. The summary shows the availability intervals with selected percentages (80%, 90%, 99%, ...), where the interval equals the worst availability among nodes. In addition, the summary shows the occurrence time of longest time between receptions for each node. This information eases finding and examining problematic times with other diagnostics. The availability chart shows the best (maximum), average, and worst (minimum) availabilities among nodes.

The logical and actual topologies of the network are shown in Figure 35. Logical topology is drawn based on the connectivity information, while the actual topology uses node locations that were configured manually to a database. The self-diagnostics allows showing usage (throughput), transmission reliability, and signal strength information per link.

The self-diagnostics allows per node, per hop, and per traffic class visualization of the end-to-end latency. Figure 36 shows the per node and per traffic class latencies in the test network. The latency information was collected by requesting latency diagnostics via normal (maximize energy weight $w_e$) and delay optimized (maximize delay weight $w_d$) traffic classes. In general, latency increases linearly as the function of hops. The relatively larger increment with 9 hops was due to momentary network

![Fig. 34. Availability summary summarizing availabilities and showing the occurrence of the largest unavailability per node (left), and availability chart with average, minimum, and maximum availabilities (right).]
errors: the largest hop count was used only after a link break that caused delays. The per traffic class results show that the latency optimized traffic class has 4 s (16%) smaller latency at 90% reception mark. Considering that the nearly linear topology severely limits available route choices, the result confirms the QoS selection via traffic classes.

Figure 37 shows received signal strengths on a selected node. The signal strength varies on short term basis although the network topology is static. The signal degradation on a next hop might break the connection necessitating time consuming neigh-

**Fig. 35.** Network topology diagnostics: logical network topology showing link reliabilities (left) and actual network topology showing link signal strengths (right).

**Fig. 36.** End-to-end latency categorized as per hop (left) and per traffic class (right).
7.6 Performance Comparison of Indoor and Outdoor Deployments

This section presents the measured network QoS in the outdoor rural area and in the indoor campus area deployments. The analysis contains network diagnostics from selected one week period. The results are summarized in Table 16.

In the indoor network, the node count varies because some of the nodes were given to students to carry around and were not therefore always within the network area. The presented node count was calculated from the number of active nodes within one hour. The area coverage and node density were calculated with known node placements, whereas the communication range is the measured range in an open space. To offload the higher traffic of the indoor network, the indoor network contained 9 sinks. This way, the network load was balanced among the sinks via the throughput component in the routing cost.

The end-to-end latency in the indoor network is smaller due to shorter hop count. Still, the average latency per hop is relatively high because the reliability of the indoor network suffered from several WLANs operating on the same frequency band. Although frequency agility methods were later developed for the TUTWSN [66], these were not in place in the examined time period.

![Connectivity analysis (223)](image)

**Fig. 37.** Link signal strength diagnostics of a selected node.
### Table 16. Comparison of QoS in the outdoor and indoor deployments.

<table>
<thead>
<tr>
<th>Performance metric</th>
<th>Outdoor</th>
<th>Indoor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Avg</td>
</tr>
<tr>
<td>Latency (s)</td>
<td>3.6</td>
<td>19.4</td>
</tr>
<tr>
<td>Hops</td>
<td>1</td>
<td>4.5</td>
</tr>
<tr>
<td>Latency/hop (s)</td>
<td>3.6</td>
<td>4.2</td>
</tr>
<tr>
<td>Throughput (bps)</td>
<td>237</td>
<td>248</td>
</tr>
<tr>
<td>Reliability (%)</td>
<td>99.78</td>
<td>99.95</td>
</tr>
<tr>
<td>Availability, 95% (s)</td>
<td>32.0</td>
<td>33.9</td>
</tr>
<tr>
<td>Lifetime (days)</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Communication range (m)</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Node count</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Area coverage (m²)</td>
<td>-</td>
<td>124000</td>
</tr>
<tr>
<td>Node density (1/1000m²)</td>
<td>0.14</td>
<td>0.14</td>
</tr>
</tbody>
</table>

As the indoor deployment has been active several years, the lifetime was calculated from the battery voltage as an elapsed time between full charge to an empty battery. However, this method was not feasible for the outdoor network due to its short, few months deployment time. Instead, its lifetime was estimated by measuring average current (500 uA) over 10 minutes and applying this to the estimated the battery capacity of 3000 mAh. It should be noted that while the measured node was part of a network, the actual lifetime may vary due to different traffic loads and reliability problems causing e.g. retransmissions.
8. SUMMARY OF PUBLICATIONS

The publications of this Thesis are based on the work of the author during years between 2005 and 2010. This chapter summarizes the contents of the publications and clarifies the contribution of the author. The co-authors agree with the described contributions of the Author. In each publication, the supervisor Prof. Marko Hännikäinen has given ideas for the designs, analyses, and experiments, and revised the draft versions of the publications. Prof. Timo D. Hämäläinen has given ideas for the publications and revised the draft versions of publications. None of the publications have previously been used as a part of a doctoral thesis.

Publication [P1] proposes a cost-based dynamic routing protocol for resource constrained WSNs. The publication presents a reactive, energy-efficient protocol messaging and QoS based route selection metrics. The protocol is verified with experimental measurements.

The author is the main architect of the routing protocol and designed the cost metrics for route selection. Mauri Kuorilehto gave ideas for the protocol and revised the text.

Publication [P2] proposes a reserved slot allocation algorithm for synchronized MACs. The algorithm adjusts slots dynamically based on traffic demands, and allows trade-off between delay and energy-efficiency determined by application and routing requirements. The algorithm is verified by modeling its operation on IEEE 802.15.4.

The author designed and analyzed the capacity optimization algorithm. Mikko Kohvakka developed the original analytical models for IEEE 802.15.4 which the author modified and extended for the performance analysis. Mauri Kuorilehto gave ideas for the algorithm.

Publication [P3] presents an embedded software architecture of WSN diagnostics. It proposes a methodology and metrics to measure QoS and determine performance problems. Also, the publication defines a configuration process for adjusting network settings to match the performance requirements.

The author designed and implemented the WSN diagnostics, while Mikko Kohvakka designed the prototype hardware used in the measurements.
Publication [P4] defines availability and reliability metrics for analyzing WSN QoS. The publication evaluates the factors affecting end-to-end QoS and the performance of the low duty cycling and beacon synchronization. Based on the results, a reliable data forwarding algorithm that guarantees end-to-end reliability is proposed for resource constrained WSNs.

The author designed and analyzed the presented metrics and protocols.

Publication [P5] presents a CoS add-on layer for wireless mesh networks. The protocol implements a priority and reservation based traffic control over contention-based MACs. It is verified with simulations and measurements on IEEE 802.11.

The author designed and implemented the CoS add-on layer, and performed the simulations and measurements.

Publication [P6] extends the reserved slot allocation algorithm by analyzing different allocation approaches. In addition, the publication describes TUTWSN MAC in detail and analyzes its performance against other low energy WSN MAC proposals.

The author designed QoS-related aspects of the TUTWSN MAC protocol comprising queuing disciplines and the slot allocation method, whereas Mikko Kohvakka was the main architect of the TUTWSN MAC layer. The author designed and wrote channel access specific parts of the publication (Sections 4.2-4.3), verified models and the performance of slot allocation methods with simulations (Section 6), and analyzed the performance of the MAC protocol in a deployment (Section 7). Mikko Kohvakka wrote the related work, generic MAC description, and designed the performance models.
9. CONCLUSIONS

The application space of WSNs is huge and each application has its distinct service requirements and characteristics. Despite the advancements in low-power communication and computing circuits, the current technology has trade-offs between available energy, physical size, computing, communication, and memory resources. Energy can be saved by reducing activity, but this necessitates QoS support in communication protocols and applications to ensure that the application requirements are fulfilled. The main research challenge is the design of scalable and adaptive QoS control for resource constrained hardware with the limited energy budget.

This Thesis presented a survey of QoS support in existing protocols and standards. The protocols were targeted for a specific purpose and did not support adjustable QoS or applications having different service requirements within the same network. Furthermore, QoS measurement and control methods concentrated only on ensuring consistent sensor values but did not consider the network operation. This motivated the design of QoS definitions, diagnostics and management tools, and QoS communication protocols.

This Thesis defined QoS that is suitable for resource constrained, low energy WSNs. The definition includes new availability metric and the selection of other relevant QoS metrics. Next, protocol designs at MAC and routing layers were developed based on the defined metrics. Instead of optimizing communication protocols for a certain use case as in the related research, the designs allow configurable and adaptive level of service. A cross-layer design was utilized to combine the MAC and routing protocols with the QoS metrics. The protocols and diagnostics were verified with simulations and practical implementation on TUTWSN platform. Finally, sensor self-diagnostics and diagnostics tools were designed to analyze and adjust the level of service in a network. Compared to the related proposals on in-network sensor diagnostics, the framework also detects performance problems and identifies reasons for the issues thus easing the repair.

The results of this Thesis can be used in the WSN research, development, and implementation in general. The presented methods and protocol designs can be adapted to
existing applications and protocols. To fulfill the vision of disposable sensor nodes, future advancements in the manufacturing technologies aim toward reduced size and cost instead of performance. While the manufacturing technologies are constantly improving, advancements that reduce cost, size, and energy at the same time are slower. Therefore, the resource constraints exist also in the near future and the principles and results of this Thesis remain valid.
BIBLIOGRAPHY


[70] IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPAN), Sep. 2006, IEEE Std 802.15.4-2006.


