Computer Communications and Networks

Dong-Seong Kim
Hoa Tran-Dang

Industrial Sensors and Controls in Communication Networks

From Wired Technologies to Cloud Computing and the Internet of Things

Springer

Computer Communications and Networks

The **Computer Communications and Networks** series is a range of textbooks, monographs and handbooks. It sets out to provide students, researchers, and non-specialists alike with a sure grounding in current knowledge, together with comprehensible access to the latest developments in computer communications and networking.

Emphasis is placed on clear and explanatory styles that support a tutorial approach, so that even the most complex of topics is presented in a lucid and intelligible manner.

More information about this series at http://www.springer.com/series/4198

Dong-Seong Kim · Hoa Tran-Dang

Industrial Sensors and Controls in Communication Networks

From Wired Technologies to Cloud Computing and the Internet of Things

Springer

Dong-Seong Kim
Department of ICT Convergence Engineering
Kumoh National Institute of Technology
Gumi, Korea (Republic of)

Hoa Tran-Dang
Department of ICT Convergence Engineering
Kumoh National Institute of Technology
Gumi, Korea (Republic of)

ISSN 1617-7975 ISSN 2197-8433 (electronic)
Computer Communications and Networks
ISBN 978-3-030-04926-3 ISBN 978-3-030-04927-0 (eBook)
https://doi.org/10.1007/978-3-030-04927-0

Library of Congress Control Number: 2018962922

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Industrial networks have been promoted increasingly by emerging technologies such as industrial wireless communication technologies, industrial Internet of Things (IIoT), cloud computing, big data, etc. Given the increasing age of many industrial distributed systems and the dynamic industrial manufacturing market, intelligent and low-cost industrial automation systems are required to improve the productivity and efficiency of such systems. The collaborative nature of industrial wireless sensor networks (IWSNs) brings several advantages over traditional wired industrial monitoring and control systems, including self-organization, rapid deployment, flexibility, and inherent intelligent-processing capability. In this regard, IWSNs play a vital role in creating a highly reliable and self-healing industrial system that rapidly responds to real-time events with appropriate actions. At broader scale, IIoT has been recognized primarily as a solution to improve operational efficiency.

In this book, detailed reviews about the emerging and already deployed industrial sensor and control network applications and technologies are discussed and presented. In addition, technical challenges and design objectives are described. Particularly, fieldbus technologies, wireless communication technologies, network architectures, resource managements, and optimization for industrial networks are discussed. Furthermore, industrial communication standards including wired and wireless technologies and IIoT visions are presented in detail. Overall, this book covers the current state of the art in such emerging technologies and discusses future research directions in this field. The book is structured in three parts, each one grouping a number of chapters describing our state-of-the-art researches in actual domains of the technology transformation in sensing and control in future industrial networks.

Part I titled as Industrial Control Networks includes six research proposals covering the fieldbus control networks (i.e., CAN, FlexRay, Modbus). In this part, the latest fieldbus technologies are reviewed to point out the key performance and challenges of technology application in industrial domain. This challenges open potential researches to find out breakthrough solutions. One of them is described in our research proposal, which proposed to use dual fieldbus technology, CAN and

Modbus to meet the significant time delay for the distributed control system of ship engines.

Part II of the book referred as Industrial Wireless Sensor Networks includes 11 research proposals which analyze and evaluate such networks' applications in terms of wireless networking performances. Such aspect is highlighted by key points composed of medium access control (MAC) mechanisms, wireless communication standards for industrial field. In additions, applications of such networks from environmental sensing, condition monitoring, and process automation applications are specified. Designing appropriate networks are based on the specific requirements of applications. It points out the technological challenges of deploying WSNs in the industrial environment as well as proposed solutions to the issues. An extensive list of IWSN commercial solutions and service providers are provided and future trends in the field of IWSNs are summarized.

Part III named as Industrial Internet of Things mentions the state-of-the-art technologies along with accompany challenges to realize such vision. Wide applications of IIoT are summarized in industrial domains. Specially, adopting such technology to the Physical Internet, an emerging logistics paradigm is described in this part.

Gumi, Korea (Republic of) Dong-Seong Kim
Hoa Tran-Dang

Acknowledgements

We would like to express our gratitude to the publisher, Springer, for accepting this book as part of the series "Computer Communications and Networks". In particular, we would like to thank Scientific Publishing Services (P) Ltd. in Tamil Nadu, India for their editorial and production guidance.

Our acknowledgment goes to Prof. Pactrick Charpentier and Dr. Nicolas Krommenacker at the Research Center for Automatic Control of Nancy (CRAN), University of Lorraine, France, who contributed to the content of Chap. 18. Special thank is due to our colleague, Nguyen Bach Long at University of Newcastle, Australia, who collaborated and contributed to the content of Chap. 17.

We would like to thank the Networked System Laboratory (http://nsl.kumoh.ac.kr/), Kumoh National Institute of Technology (KIT) in Gumi, Republic of Korea, which made it possible to write this book. As one of the most educational institutions in technology, we are grateful to be part of this visionary and innovative organization.

Contents

Part I Industrial Control Networks

1 An Overview on Industrial Control Networks 3
1.1 Introduction 3
1.2 Architecture of Industrial Control Networks 4
1.3 Requirements of Industrial Control Networks 6
1.4 Communication Technologies for Industrial Control Networks 8
1.4.1 Fieldbuses 8
1.4.2 Industrial Ethernet 11
1.5 Trends and Issues 14
1.6 Conclusions 15
References 16

2 FlexRay Protocol: Objectives and Features 17
2.1 Introduction 17
2.2 FlexRay System 18
2.2.1 Level 1—Network Topology 18
2.2.2 Level 2—Interface 18
2.2.3 Level 3—CHI and Protocol Engine 19
2.3 Message Scheduling for FlexRay System 20
2.3.1 FlexRay Static Segment 20
2.3.2 FlexRay Dynamic Segment 23
2.3.3 Comparison with CAN 24
2.4 Verification and Validation 25
2.4.1 Computer Simulation for Model Validation 25
2.4.2 Formal Verification 26
2.5 Software and Hardware 27
2.5.1 Software 27
2.5.2 Hardware 27

2.6 Conclusions . . . 29
References . . . 29

3 Communication Using Controller Area Network Protocol . . . 31
3.1 Introduction . . . 31
3.2 CAN Protocol Overview . . . 33
3.2.1 Physical Layer . . . 33
3.2.2 Message Frame Format . . . 34
3.2.3 Medium Access Technique . . . 36
3.2.4 Error Management . . . 38
3.2.5 Implementation . . . 39
3.3 Main Features . . . 39
3.3.1 Advantages . . . 39
3.3.2 Performances . . . 40
3.3.3 Determinism . . . 40
3.3.4 Dependability . . . 41
3.4 Conclusions . . . 41
References . . . 41

4 Distributed Control System for Ship Engines Using Dual Fieldbus . . . 43
4.1 Introduction . . . 43
4.2 Redundant Distributed Control System . . . 46
4.2.1 Modbus Protocol . . . 49
4.2.2 CAN Protocol . . . 51
4.2.3 Redundancy . . . 53
4.3 Implementation and Experimental Test . . . 56
4.4 Conclusions . . . 62
References . . . 62

5 Implementing Modbus and CAN Bus Protocol Conversion Interface . . . 65
5.1 Introduction . . . 65
5.2 Modbus and CAN Bus . . . 66
5.2.1 Modbus . . . 66
5.2.2 CAN Bus . . . 68
5.3 Conversion Interface Design . . . 70
5.3.1 Hardware Design . . . 70
5.3.2 Software Design . . . 71
5.4 Conclusions . . . 72
References . . . 72

6 MIL-STD-1553 Protocol in High Data Rate Applications . . . 73
6.1 Introduction . . . 73
6.2 Related Works . . . 74

6.3 MIL-STD-1553 Network Protocol Infrastructure 75
6.3.1 MIL-STD-1553 Hardware Elements 75
6.3.2 MIL-STD-1553 Protocol Format 78
6.3.3 Manchester Encoder/Decoder 78
6.3.4 Quality Control Process 79
6.4 Comparative Analysis of High-Speed Data Bus Technologies 81
6.4.1 Traditional MIL-STD-1553 Architecture 81
6.4.2 HyPer-1553TM Data Bus Technology 81
6.4.3 Turbo 1553 Approach 83
6.4.4 Tools for Testing and Simulation 85
6.5 Conclusions and Future Works 86
References 87

7 Research and Design of 1553B Protocol Bus Control Unit 89
7.1 Introduction 89
7.2 1553B Protocol 90
7.2.1 Hardware Characteristics 90
7.2.2 Encoding 90
7.2.3 Word and Message 90
7.2.4 Hierarchical Division 91
7.3 BCU Design 91
7.3.1 Decoding Unit 92
7.3.2 Data Encode Unit 94
7.3.3 Command Words Decode Unit 94
7.3.4 Send Control Unit 94
7.3.5 Status Words Receive Control and Decode Unit 95
7.3.6 Address Decode Unit 95
7.3.7 Send Overtime Detection Unit 95
7.3.8 Error Detection Unit 96
7.3.9 DSP Communication Interface 96
7.4 Logic Emulation 96
7.5 Conclusions 97
References 97

Part II Industrial Wireless Sensor Networks

8 An Overview on Wireless Sensor Networks 101
8.1 Introduction 101
8.2 Wireless Sensor Networks 102
8.3 Network Topologies of Wireless Sensor Networks 103
8.4 Applications of WSNs 105
8.4.1 Application Classification 106
8.4.2 Examples of Application Requirements 107

8.5 Characteristic Features of Wireless Sensor Networks 109
8.5.1 Lifetime 109
8.5.2 Flexibility 110
8.5.3 Maintenance 110
8.6 Existing Technologies and Applications 111
8.7 Conclusions 112
References 113

9 Wireless Fieldbus for Industrial Networks 115
9.1 Introduction 115
9.2 Wireless Fieldbus Technology 118
9.2.1 Overview 118
9.2.2 Wireless Fieldbus Systems Proposals 119
9.3 Issues in Wireless Fieldbus Networks 122
9.3.1 Consistency Problems of Fieldbus Technology 123
9.3.2 Problems for Token-Passing Protocols 123
9.3.3 Problems in CSMA Based Protocol 124
9.4 Conclusions 124
References 124

10 Wireless Sensor Networks for Industrial Applications 127
10.1 Introduction 127
10.2 Industrial Wireless Sensor Networks 129
10.2.1 Safety Systems 129
10.2.2 Closed-Loop Regulatory Systems 129
10.2.3 Closed-Loop Supervisory Systems 130
10.2.4 Open Loop Control Systems 130
10.2.5 Alerting Systems 130
10.2.6 Information Gathering Systems 130
10.3 Industrial Standards 130
10.3.1 ZigBee 131
10.3.2 WirelessHART 132
10.3.3 ISA100.11a 134
10.4 Wireless Sensor Networks for Industrial Applications 136
10.4.1 Industrial Mobile Robots 137
10.4.2 Real-Time Inventory Management 137
10.4.3 Process and Equipment Monitoring 138
10.4.4 Environment Monitoring 139
10.5 Conclusions 139
References 140

11 MAC Protocols for Energy-Efficient Wireless Sensor Networks . . . 141
11.1 Introduction 141
11.2 MAC Layer-Related Sensor Network Properties 142

11.2.1 Reasons of Energy Waste . . . 142
11.2.2 Communication Patterns . . . 142
11.2.3 Properties of a Well-Defined MAC Protocol . . . 143
11.3 Multiple-Access Consideration in Sensor Network Properties . . . 143
11.3.1 Network Topologies . . . 144
11.3.2 Time-Division Multiple Access (TDMA) . . . 146
11.3.3 Carrier-Sense Multiple Access (CSMA) and ALOHA . . . 147
11.3.4 Frequency-Division Multiple Access (FDMA) . . . 148
11.3.5 Code-Division Multiple Access (CDMA) . . . 148
11.4 Proposed MAC Layer Protocols . . . 149
11.4.1 Sensor-MAC . . . 149
11.4.2 WiseMAC . . . 150
11.4.3 Traffic-Adaptive MAC Protocol . . . 152
11.4.4 Sift . . . 153
11.4.5 DMAC . . . 153
11.4.6 Timeout-MAC/Dynamic Sensor-MAC . . . 154
11.4.7 Integration of MAC with Other Layers . . . 155
11.5 Open Issues and Conclusion . . . 156
References . . . 158

12 Cooperative Multi-channel Access for Industrial Wireless Networks Based 802.11 Standard . . . 161
12.1 Introduction . . . 161
12.2 Throughput Enhancement . . . 162
12.2.1 CAMMAC-802.11 . . . 162
12.2.2 Using Directional Antennas . . . 164
12.2.3 Negotiation-Based Throughput Maximization Algorithm . . . 165
12.3 Access Delay . . . 167
12.4 Mitigating the Impact of Inter-node Interference . . . 170
12.5 Conclusions . . . 171
References . . . 172

13 802.11 Medium Access Control DCF and PCF: Performance Comparison . . . 173
13.1 Introduction . . . 173
13.2 IEEE 802.11 Media Access Protocols . . . 174
13.2.1 Distributed Coordinate Function (DCF) . . . 175
13.2.2 Point Coordinate Function (PCF) . . . 176
13.3 Performance Comparison . . . 177
13.4 Conclusions . . . 178
References . . . 179

14 An Overview of Ultra-Wideband Technology and Its Applications 181
14.1 Introduction 181
14.2 History and Background 182
14.3 UWB Concepts 183
14.3.1 High Data Rate 184
14.3.2 Low Power Consumption 185
14.3.3 Interference Immunity 185
14.3.4 High Security 185
14.3.5 Reasonable Range 185
14.3.6 Large Channel Capacity 185
14.3.7 Low Complexity, Low Cost 186
14.3.8 Resistance to Jamming 186
14.3.9 Scalability 186
14.4 UWB Technologies 187
14.4.1 Impulse Radio 187
14.4.2 Multiband OFDM 187
14.4.3 Comparison of UWB Technologies 189
14.5 Technologies and Standards 189
14.5.1 Bluetooth 189
14.5.2 UWB 190
14.5.3 UWB Standards 191
14.5.4 Marketplace and Vendor Strategies 193
14.6 UWB Applications 193
14.6.1 Communications 194
14.6.2 Radars/Sensors 195
14.7 Conclusions 195
References 196

15 Ultra-Wideband Technology for Military Applications 197
15.1 Introduction 197
15.2 Technical Overview of Ultra-Wideband Systems 198
15.3 Ultra-Wideband Technology for Military Applications 199
15.4 Conclusions 203
References 204

Part III Industrial Internet of Things

16 An Overview on Industrial Internet of Things 207
16.1 Introduction 207
16.2 Architecture of IIoT System 207
16.3 Key Enabling Technologies for IIoT 210
16.3.1 Identification Technology 210
16.3.2 Sensor 211

16.3.3 Communication Technology . . . 211
16.3.4 IIoT Data Management . . . 212
16.3.5 Cloud Computing . . . 212
16.4 Major Application of IIoT . . . 213
16.4.1 Heath Care . . . 213
16.4.2 Logistics and Supply Chain . . . 213
16.4.3 Smart Cities . . . 213
16.5 Conclusions . . . 215
References . . . 215

17 Energy-Aware Real-Time Routing for Large-Scale Industrial Internet of Things . . . 217
17.1 Introduction . . . 217
17.2 Related Works . . . 220
17.3 System Model . . . 222
17.3.1 Network Topology . . . 222
17.3.2 Variable Definition . . . 223
17.3.3 Energy Model . . . 223
17.4 Energy-Aware Real-Time Routing Scheme (ERRS) . . . 225
17.4.1 Clustering Scheme . . . 225
17.4.2 Routing Scheme . . . 228
17.5 Performance Evaluation . . . 230
17.5.1 IEEE 802.15.4a CSMA/CA Scheme for IIoT . . . 230
17.5.2 Simulation Model . . . 231
17.5.3 Simulation Results . . . 232
17.6 Conclusions . . . 237
References . . . 237

18 3D Perception Framework for Stacked Container Layout in the Physical Internet . . . 241
18.1 Introduction . . . 241
18.2 Literature Review . . . 243
18.3 Problem and Methodology . . . 244
18.3.1 Problem Definition and Proposed Approach . . . 244
18.3.2 Methodology and Assumptions . . . 246
18.4 Mathematical Formulation of the CSP Problem . . . 249
18.4.1 Parameters and Variables . . . 249
18.4.2 Formulation . . . 250
18.5 Application and Results . . . 252
18.5.1 Experimental Setup . . . 252
18.5.2 Results and Discussions . . . 253
18.6 Conclusion and Future Works . . . 256
References . . . 258

19 An Information Framework of Internet of Things Services for Physical Internet 259
19.1 Introduction 259
19.2 IOT Infrastructure for Physical Internet 263
19.2.1 π-Containers 263
19.2.2 π-Movers 266
19.2.3 π-Nodes 267
19.2.4 Active Distributed PIMS for π-Nodes 268
19.3 Service-Oriented Architecture for the IOT 272
19.3.1 Physical Layer 273
19.3.2 Network Layer 274
19.3.3 Service Layer 274
19.3.4 Interface Layer 275
19.4 Management of Composite π-Containers: A Case Study 275
19.4.1 Architecture 276
19.4.2 An Information Flow Framework to Retrieve 3D Layouts 278
19.4.3 Value-Added Services Enabled by Retrieved 3D Layouts 279
19.5 Conclusion and Future Works 279
References 280

Index 283

Part I
Industrial Control Networks

Industrial control networks play a significant role in industrial contributed control systems since it enables all the system components to be interconnected as well as monitor and control the physical equipment in industrial environments. With the development of electronic engineering, the mechanical control system has been gradually replaced by digital control systems adopting power microprocessors and digital controllers. The movement toward such digital systems requires inherently corresponding communication technologies and communication protocols to the field as well as the controllers. Basically, the core of industrial networking consists of fieldbus protocols which are defined in the IEC standard 61158 as a digital serial, multi-drop, data bus for communication with industrial control and instrumentation devices.

Part I of this book named Industrial Control Networks includes six research proposals covering the key fieldbus control networks (i.e., CAN, FlexRay, and Modbus). In this part, such fieldbus technologies are reviewed to point out the key performance and challenges when applying such communication technologies in industrial control systems. The challenges open up potential researches to find out breakthrough solutions aiming to improve the quality of services of the systems. One of them is described in our research proposal, which proposed to use dual fieldbus technology, CAN, and Modbus to meet the significant time delay for the distributed control system of ship engines.

Chapter 1
An Overview on Industrial Control Networks

1.1 Introduction

In general, the industry can be divided into two categories: process, and manufacturing sector. The process industry deals with processes involving very large material flows in both continuous, or discontinuous manner and often has strict safety requirements (e.g., power generation, cement kilns, petrochemical production), while the manufacturing industry is concerned with the production of discrete objects. Achieving the maximum throughput of produced goods is, normally, very important aspect in such industrial sectors. Practically, the industrial systems have been required to be innovated to enhance production monitoring and quality control and in the same time maintaining the operation costs as low as possible. This innovation has happened in the last few decades due to the advancement of information and communication technologies which enable the industrial systems to match up with these needs. The innovation has led to reduce significantly manual labors replaced with a faster, and more reliable automated machine, equipment in the most of industry operations. This also provides both the factories and the manufacturing plants with necessary monitoring which they both sought for better supervisory and quality control. Introducing all this number of automated unites into the factories needed an efficient method to connect them together, to communicates with each other, and to transfer the various supervisory data to the monitors. This leads to the introduction of the communication networks into the industrial sectors.

Based on the specialized functions, the industrial networks are composed of three major control components that include Programmable Logic Controllers (PLC), Supervisory Control and Data Acquisition (SCADA), and Distributed Control Systems (DCS) [1]. PLCs are nothing but digital computers that can work in hazardous industrial environments. Such processor-based systems take inputs from data generation devices like sensors and communicate them with the entire production unit and then present the output to HMI (Human–Machine Interfaces). PLCs can control the entire manufacturing process while ensuring the required quality of services

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_1

(QoS) and great precision control functions. SCADA systems are mainly used for the implementation of monitoring and control system of an equipment or a plant in several industries like power plants, oil and gas refining, water and waste control, telecommunications, etc. In this system, measurements are made under field or process level in a plant by number of remote terminal units and then data are transferred to the SCADA central host computer so that more complete process or manufacturing information can be provided remotely. This system displays the received data on number of HMIs and conveys back the necessary control actions to the remote terminal units in process plant. DCS consists of a large number of local controllers in various sections of plant control area and are connected via a high-speed communication network. In DCS control system, data acquisition, and control functions are carried through a number of DCS controllers which are microprocessor-based units distributed functionally and geographically over the plant and are situated near area where control or data gathering functions being performed.

All these three elements deals with field instruments (i.e., sensors, actuators), smart field devices, supervisory control PCs, distributed I/O controllers and HMI (Human–Machine Interface). These devices are connected and communicated by a powerful and effective communication network referred to as industrial networks. In these networks, the data or control signals are transmitted either by wired or wireless media. Cables used for wired transmission of data include twisted pair, coaxial cable or fiber optics. Meanwhile, radio waves are used to transmit data in the industrial wireless networks.

1.2 Architecture of Industrial Control Networks

Generally, the industrial control networks are constructed in hierarchical topology as illustrated in Fig. 1.1 including three basic levels: informational, control, device level [2]. Each level has unique requirements that affect which network is used for that particular level.

The device level consists of field devices such as sensors and actuators of processes and machines. The task of this level is to transfer the information between these devices and technical process elements such as PLCs. The information transfer can be digital, analog, or hybrid. The measured values may stay for longer periods or over a short period. All of the devices connect to a single cable. The cable usually has conductors for power, device signal, and a shield. There are many other field level communication networks available which are characterized by different factors such as response time, message size, etc. The messages are usually small when compared to other networks. Because of deterministic and repeatability requirements, the messages can be prioritized so that the more critical information is transmitted first [3]. Nowadays, fieldbus technology is the most sophisticated communication network used in field level as it facilitates distributed control among various smart field devices and controller. These networks support Carrier-Sense Multiple

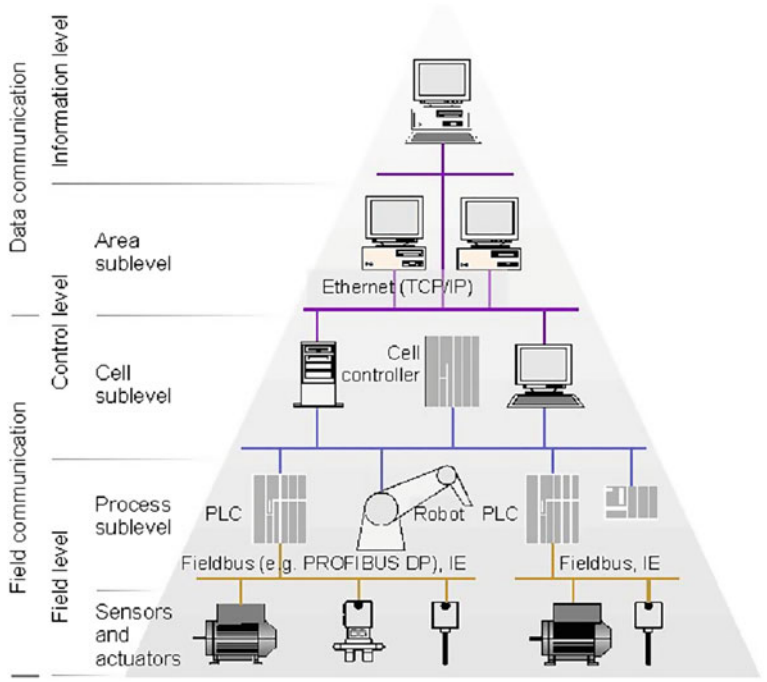


Fig. 1.1 Three-level architecture of industrial control networks

Access with Arbitration on Message Priority (CSMA/AMP) protocol for fulfilling the requirements.

The control level involves networking machines, work cells, and work areas. This is the level where Supervisory Control and Data Acquisition (SCADA) is implemented. If an automotive assembly plant is used as an example, this network level is where the individual control systems will be given information on the make, model, and options that are to be included on a vehicle so that the controllers can run the appropriate programs to assemble the vehicle correctly. Data such as cycle times, temperatures, pressures, volumes, etc., are also collected at this level [4]. The tasks of this level include configuring automation devices, loading of program data and process variables data, adjusting set variables, supervising control, displaying variables data on HMIs, historical archiving, etc. The control level of the network must achieve the predefined requirements such as deterministic, repeatable, short response time, high-speed transmission, short data lengths, machine synchronization, and constant use of critical data. Determinism is the ability to accurately predict when data will be delivered, and repeatability is the ability to ensure that transmit times are consistent and unaffected by devices connecting to the network. Because of the deterministic and repeatability constraints, medium access control protocol by CSMA/CD (Carrier-Sense Multiple Access with Collision Detection) in traditional Ethernet network is inadequate, so a different type of network access is required. Local Area Networks

(LANs) are widely used as communication networks in this level to achieve desired characteristics. The Ethernet with TCP/IP protocol is mostly used as a control level network to connect control units with computers. In addition, this network acts as a control bus to coordinate and synchronize between various controller units. Some fieldbuses are also used in this level as control buses such as PROFIBUS and ControlNet. In such networks, Concurrent Time Domain Multiple Access (CTDMA) protocol is used based on a time-slice algorithm that regulates each node's opportunity to transmit in a network interval. Adjustment of the amount of time for a network interval gives a consistent, predictable time for data transmission [3].

The informational level is the top level of the industrial system which gathers the information from the lower level, i.e., control level. It deals with large volumes of data that are neither in constant use or time critical. Large-scale networks exist in this level. So Ethernet WANs are commonly used as information level networks for factory planning and management information exchange. Sometimes these networks may connect to other industrial networks via gateways.

1.3 Requirements of Industrial Control Networks

The industrial control networks operate along two different paradigms: time-triggered and event-triggered [5]. In the time-triggered applications, the systems work periodically. They first wait for the beginning of the period (or some offset from the beginning), sample their inputs, and compute some algorithm according to the inputs and some set point data received from computers higher in the hierarchy. They then make the results available at the outputs. Inputs and outputs correspond to sensors and actuators at the lowest level in the hierarchy. At higher levels, inputs correspond to status and completion reports from the next lower level. Outputs are set points or commands to the lower level. Acquisition and distribution applications are special cases. Acquisition applications have no outputs to the process but store the output results internally. Distribution applications have no input and compute the algorithms from stored information.

Periodicity is not mandatory but often assumed because it simplifies the algorithms. For instance, most digital control theory assumes periodicity. Furthermore, it assumes limited jitter on the period and bounded latency from input instant to output instant. Acquisition and distribution applications have similar requirements in terms of periodicity and jitter.

Meanwhile, in event-triggered applications, the system is activated upon the occurrence of events. An event may be the arrival of a message with a new command or a completion status or the change of an input detected by some circuitry. When an event is received, the application computes some algorithm to determine the appropriate answer. The answer is then sent as an event to another application locally or remotely. The time elapsed between the generation of the input event and the reception of the corresponding answer must be bounded. Its value is part of the requirements on the application and also the communication system if the events have to be transported

through some network. Furthermore, applications should be able to assess some order in the event occurrences. This is usually not a problem when the event is detected and processed on the same computer. It becomes a problem when the events are detected at different locations linked by a network which may introduce some variable delay.

Because large amounts of data may be passed through these control networks and message lengths tend to be longer, data transmission rates tend to be faster than with fieldbus networks. However, since they can be used to pass time-critical data between controllers, control networks must also be deterministic and meet the time-dependent (usually called real time) needs of their intended applications. Determinism in a network context is defined as: there is a specified worst-case delay between the sensing of a data item and its delivery to the controlling device. Real time in this context is defined as "sufficiently rapid to achieve the objectives of the application," and is measured as latency time. Determinism and latency are separate but complementary requirements. Both determinism and a specific latency are required to achieve synchronization. If the same control network is used to exchange both real-time data between controllers and business information between controllers and business systems, clearly there must be some way to prevent business information from interfering with real-time deterministic response. Many complex protocols have been constructed for this purpose, but most control networks rely only on the underlying nature of the chosen network protocol. Usually, determinism is achieved by preventing message collisions and limiting the maximum message length. Low latency is achieved by using high-speed media and minimizing the number of times a signal must be rebroadcast, such as in a mesh network.

The benefit of using a standard network protocol such as Transport Control Protocol/Internet Protocol (TCP/IP) over an Ethernet network is a lower cost. By simply selecting standard Ethernet cabling and using full-duplex Ethernet switches instead of passive hubs, a control network built on this commodity technology can guarantee that there will be no network collisions, making such networks deterministic. Using high speed such as 100 or 1000 Mbps and the standard Ethernet maximum packet length of 1500 bytes achieves low latency and means that other applications cannot "hog the wire," preventing time-critical data transfers. However, you still must do the math! Remember that the definition of real time and determinism requires that the network must make its bandwidth available for time-critical data transfers in less than the maximum time period allowed for the control system. For example, if a business application were to transfer a maximum size Ethernet message (1500 bytes) at 100 Mbps, the network would be blocked for a maximum time of about 150 μs. Normally this magnitude of delay is perfectly acceptable for both process control and factory automation needs, but it may not be acceptable for motion control or machine control. It would be nice if control networks and fieldbus networks could not be used for the same applications, but they can. It would also be nice if control networks were always confined to a business or control room environment, but increasingly they are being extended to the field and shop floor. In some cases, control networks

are being used in applications normally requiring a fieldbus. In fact, all of the control networks were developed from one or more of the fieldbus networks and use the same application layer and user layer protocols. Since control networks are related to fieldbuses, there will continue to be a very loose dividing line between them.

1.4 Communication Technologies for Industrial Control Networks

In order to carry out the assigned tasks of networks, it is essential for the devices to communicate. Traditional wired communication technologies have played a crucial role in industrial monitoring and control networks. Accordingly, this communication was performed over point-to-point wired systems. Such systems, however, involved a huge amount of wiring which in turn introduced a large number of physical points of failure, such as connectors and wire harnesses, resulting in a highly unreliable system. These drawbacks resulted in the replacement of point-to-point systems using advanced industrial communication technologies. This section aims to highlight the major technologies deployed in the industrial control networks.

1.4.1 Fieldbuses

Traditionally, control systems in factories and plants were analog in nature. They used direct connections from controller to actuator or transducer to controller and were based on a 4- to 20-mA control signal. As the system became more complex and as networking technologies evolved, eventually a change from the analog system to a digital system came about. Fieldbuses were developed to tie all these digital components together. Over the past few decades, the industry has developed a myriad of fieldbus protocols (e.g., Foundation Fieldbus H1, ControlNet, PROFIBUS, CAN, etc.). Compared to traditional point-to-point systems, fieldbuses allow higher reliability and visibility and also enable capabilities, such as distributed control, diagnostics, safety, and device interoperability [6].

Fieldbuses are digital networks and protocols that are designed to replace the analog systems. They are essentially industrial LANs that network all of the computers, controllers, sensors, actuators, and other devices so they can interact with one another. A single network cable replaced the dozens or even hundreds of individual analog cables in the older systems. Protocols allowed operators to easily monitor, control, troubleshoot, diagnose, and manage all devices from a central location. While these fieldbuses reduced the wiring and improved the reliability and flexibility of the system, another issue was created: multiple proprietary systems, with incompatibility and a lack of interoperability between the various components. Devices made to work with one fieldbus and protocol could not work with another.

Many fieldbuses have been developed. Some attempts at building a common standard were made but no one system or standard ever emerged. ControlNet, DeviceNet, Foundation Fieldbus H1, HART, Modbus, PROFIBUS PA, and CAN/ CAN open are the most commonly used fieldbuses.

1.4.1.1 ControlNet

ControlNet is an industrial network and protocol supported by the Open DeviceNet Vendors Association (ODVA) [7]. It is based on the Common Industrial Protocol (CIP), which defines messages and services to be used in manufacturing automation. ControlNet uses RG-6 coax cable with Bayonet Neill-Concelman (BNC) connectors for the physical layer (PHY) and is capable of speeds to 5 Mbits/s using Manchester coding. The topology is a bus with a maximum of 99 drops possible. Its timing permits a form of determinism in the application.

1.4.1.2 DeviceNet

DeviceNet is another ODVA-supported fieldbus. It uses the well-known controller area network (CAN) technology for the PHY that was originally developed by Bosch for automotive applications. The DeviceNet protocol is similar to that of ControlNet and also uses the Common Industrial Protocol (CIP) at the upper layers. Layers 1 and 2 are CAN bus. The medium is an unshielded twisted pair (UTP) using a single-ended non-return-to-zero (NRZ) format with logic levels of 0 V and +5 V. The topology is a bus with up to 64 nodes allowed. Data rate depends on bus length and can be as high as 1 Mbit/s at 25 m to 125 kbits/s at 500 m.

1.4.1.3 Foundation Fieldbus H1

Originally developed by the International Society of Automation (ISA) standards group as Foundation Fieldbus, SP50 was one of the earlier digital fieldbuses for replacing 4- to 20-mA loops. The protocol is designated H1, which uses the IEC 61158-2 standard for the PHY and features twisted-pair cabling with basic data rate of 31.25 kbits/s. The transmission frames are synchronous with start and stop delimiters. Coding is Manchester.

1.4.1.4 HART

HART standing for Highway Addressable Remote Transducer is a two-way communications path over twisted pair. It retains the popular 4–20 mA analog functionality but adds digital signals. The digital signal is in the form of a frequency shift keying

(FSK) modulated carrier that uses the old Bell 202 modem frequencies of 2200 Hz for a 0 and 1200 Hz for a 1.

The data rate is 1200 bits/s. The FSK signal is phase continuous and does not affect the analog signal level because it is ac. Also, the FSK signal is a 1-mA variation around the dc level. The protocol uses OSI layers 1 through 4 and 7. The digital part of the communications is primarily used for commands, provisioning, and diagnostics.

The HART fieldbus is popular as it is compatible with older 4- to 20-mA equipment while adding the digital networking capability. It is still widely used. Typical HART field instruments such as the Analog Devices ADμCM360 consist of an embedded controller and the I/O for the sensors such as a pressure transducer and real-time data (RTD) temperature sensor. The on-board 24-bit sigma-delta analog-to-digital converters (ADCs) digitize the sensor information and then send it to the AD5421, a digital-to-analog converter (DAC) and 4- to 20-mA current source for connection to the cable. Digital information is also sent to the AD5700 HART FSK modem.

1.4.1.5 Modbus

Modbus is a popular industrial protocol normally used for communications with PLCs. It is simple and the standard is open so that any users can use it. Basically, Modbus works with RS-232 interfaces. The basic format comprises asynchronous characters sent and received with a UART. Modbus can be carried over a variety of PHYs and is often encapsulated in Transmission Control Protocol/IP (TCP/IP) and transmitted over Ethernet. It is also compatible with a wireless link.

1.4.1.6 PROFIBUS

PROFIBUS, another widely used fieldbus, was developed in Germany and is popular worldwide. There are versions for decentralized peripherals (DP) and process automation (PA). The protocol is synchronous and operates in OSI layers 1, 2, 4, and 7. Using RS-485, bit rates can range from 9.6 kbits/s to 12 Mbits/s. With a bus up to 1900 m long, the data rate is 31.25 kbits/s.

1.4.1.7 CAN/ CANopen

The use of Controller Area Network (CAN) is still dominated by its vast use in the automobile industry. Another stronghold is the use as a physical layer for the SAE J1939 protocol, and CAN will remain the most cost-sensitive fieldbus solution for small, embedded systems. In summary, the use of CAN will continue in:

- Automobiles and Trucks
- Aerospace (e.g., satellites)
- SAE J1939

- Small Embedded Solutions
- Legacy Applications

CANopen is basically a software add-on to provide network management function to CAN. The side effect is a reduced CAN bandwidth. These CANopen legacy applications are motion control and Industrial Machine Control.

CAN and CANopen, used as fieldbus systems for embedded solutions. The advantages of such networks include

- Extreme Reliability and Robustness
- No Message Collision
- Very low resource requirements
- Low-cost implementation
- Designed for real-time applications
- Very short error recovery time
- Support of device profiles (CANopen only)

However, there are some disadvantages of using CAN and CANopen, the biggest being the limited network length (~120 ft at a 1 Mbit/s baud rate). The disadvantages include limited network length (depending on baud rate), the limited baud rate of 1 Mbit/s, and limited bandwidth.

1.4.2 Industrial Ethernet

For many years, Controller Area Network (CAN) and CANopen, a higher layer protocol based on CAN, has been proved to be the best solution for low-cost industrial embedded networking. However, the most obvious shortcomings of these technologies include limited baud rate and limited network length. Industrial Ethernet technologies are currently the most formidable challenge to CANopen as the low-cost industrial networking technology of choice for business and enterprise for decades. It is by far the most successful and widely used networking technology in the world. It is affordable and reliable and is backed up by a strong series of IEEE 802.3 standards that keep it current. Over the past 10 years or so, Ethernet has found its way into the industrial setting for I/O and networking. It is gradually replacing the multiple fieldbuses and proprietary networks or working with them.

Some of the benefits of moving to Ethernet are

- Fewer smaller networks: Most fieldbuses can connect up to 20–40 devices. But beyond that, a separate fieldbus network is required for more devices and for connecting the two networks, if that is even possible. With Ethernet, you can connect up to 1000 devices on the same network. This arrangement improves the efficiency and decreases the complexity of the network.
- Lower cost: With many Ethernet vendors, equipment prices are competitive and the overall cost of building a network is typically lower than building a fieldbus network.

- Higher speeds: Ethernet has much higher speed capability than most fieldbuses. While that speed is not always needed, it is a benefit and the network grows in size and as faster devices are connected. While 10/100-Mbits Ethernet is the most common, some industrial facilities have already upgraded to 1-Gbit/s Ethernet.
- Connection to the factory or plant IT network: The industrial networks are traditionally kept separate from the business network, but companies are finding that advantages occur when the two networks can be interconnected. Data can be collected and used to optimize the manufacturing process or make improvements or decisions not previously possible.
- Connection to the Internet: This may not be desirable, but if it is, Ethernet provides a very convenient way to send and receive data over an Internet connection.

There are two main disadvantages of using Ethernet in the industrial setting. First, the hardware was not designed for the demanding environment in factories and process plants. Excessive temperatures, environmental hazards, chemicals, dust, mechanical stresses, and moisture make traditional equipment less reliable. Yet over the years, manufacturers have repackaged Ethernet gear to bear up under such conditions by adding industrial-grade housings and tougher electronics. Another hazard is excessive noise caused by motors, power switching, and other sources. Ethernet's differential wiring is essentially noise-resistant. It can be made noise-free with shielded cable. Most industrial wiring is simply standard, but higher grade CAT5 or CAT6 cable usually suffices. On the other hand, the RJ-45 connectors are a source of problems, especially in dirty and high-moisture environments. Special RJ-45 connectors have been developed to solve this problem. These connectors add a dirt- and moisture-sealed cover to an upgraded RJ-45 connector. These connectors meet the rigid IP67 environmental standards for hazardous environments.

A second disadvantage is Ethernet's inherent non-deterministic nature. Many industrial networks rely on timing conditions that must occur within a specific time frame. Many need a real-time connection or something close to it. Determinism means that a device or system can respond within a minimum time interval. It can respond in less time but no more than a specified time. If a device is deterministic to 10 ms, then anything less is okay. The response does not usually have to be repeatable, but that depends on the application.

Ethernet determinism is widely variable. It is a function of the carrier-sense multiple access with collision detection (CSMA/CD) access method, cable lengths, number of nodes, and the combinations of hubs, repeaters, bridges, switches, and routers used in the system. To improve the deterministic response, designers of industrial Ethernet systems must keep cables short and minimize the number of nodes, hubs, and bridges. Switches can be added to larger networks to isolate different segments, and that reduces the number of collisions and interactions.

Determinism can also be implemented or improved in some cases by using the IEEE 1588 Precision Time Protocol (PTP). The PTP permits systems with clocks to achieve synchronization among all connected devices, allowing precise timing information transfer within a network. Time stamping and near real-time performance can occur in some applications.

Another feature of Ethernet finding acceptance is Power over Ethernet (PoE). Defined by IEEE standards 802.3af and 802.3at, it allows the transmission of dc power over the Ethernet cables to power remote devices. This is a major benefit in many industrial settings and eliminates the need to install a power source near some remote sensor or other device. The standard defines power levels up to 15.4 or 25.5 W, but higher power versions up to 51 W are becoming available.

Finally, several enhanced or modified versions of Ethernet have been developed to overcome the timing issues of standard Ethernet or simply make it more compatible with existing equipment and systems. These include EtherCAT, EtherNet/IP, Profinet, Foundation Fieldbus HSE (high-speed Ethernet), and Modbus/TCP. Some use special protocols while others use TCP/IP.

EtherNet/IP is an application layer protocol using CIP, which defines all devices as objects and specifies the messages, services, and transfer methods. CIP is then encapsulated in a TCP or User Datagram Protocol (UDP) packet for transfer over Ethernet.

Profinet is another protocol that uses TCP/IP over Ethernet. It is not PROFIBUS over Ethernet. Instead, it uses two different protocols: one called Profinet CBA for component-based systems and Profinet IO for real-time I/O operations. Profinet CBA can provide determinism in the 100-ms range. It also can deliver determinism to 10 ms. A version of Profinet IO called IRT for isochronous real time can have a determinism of less than 1 ms.

Foundation Fieldbus HSE uses the H1 protocol over TCP/IP. It also uses a special scheduler that helps to guarantee messages in known times to ensure determinism at some desired level.

EtherCAT gets rid of the CSMA/CD mechanism and replaces it with a new "telegram" message packet that can be updated on the fly. Networked devices are connected in a ring or a daisy chain format that emulates a ring. As data is passed around the ring, message data can be stripped off or inserted by the addressed node while the data is streaming. The one or more EtherCAT telegrams are transported directly by the Ethernet frame or encapsulated into UDP/IP datagrams. Determinism of 30 μs and less can be achieved with up to 1000 nodes.

Modbus/TCP is the popular Modbus fieldbus protocol packaged in a TCP/IP packet. The Modbus checksum is replaced by TCP/IPs 32-bit checksum. Then the TCP/IP packets are carried over standard Ethernet.

Obviously, all of these systems are not interoperable with one another. But they can all coexist on the same Ethernet LAN since they all conform to the Ethernet Layer 1 PHY standard. Those using TCP/IP could be made interoperable with the appropriate software modifications.

Table 1.1 provides an overview of the various industrial Ethernet protocols [1].

Table 1.1 Comparison of various industrial ethernet networks

	EtherCAT	Ethetnet/IP	Powerlink	Modbus/TCP
Vendor Organization	EtherCAT Technology Group	Open DeviceNet Vendor Organization	Ethernet Powerlink Specification Group	Modbus-IDA Group
Homepage	www.ethercat.org	www.odva.org	www.ethernet-powerlink.org	www.modbus-ida.org
Availability of specification	Members signing an NDA	Free	Members	Free
Availability of technology	Example Code, ASIC, FPGA	Example Code	Standard Ethernet Chips	Example Code
Products available since	2003	2000	2001	1999
Interaction structure	Master/Slave	Client/Server	Master/Slave	Client/Server
Communication method	One frame for all communication partners	Message oriented	Message oriented	Message oriented
Ether data transfer rate	100 Mbit/s	100/10 Mbit/s	100 Mbit/s	100/10 Mbit/s
Physical topology	Line, Daisy, Chain, Tree	Star	Star	Star, Tree
Logical topology	Open Ring Bus	Bus	Ring	Bus
Infrastructure components	Switches between different segments	Switches (hubs are possible, but not efficient)	Hubs, no switches	Hubs, switches
Device profiles	CANopen, SERCOS	DeviceNet, ControlNet	CANopen	None

1.5 Trends and Issues

The industrial field generally lags behind other sectors of electrics simply because its technological needs do not follow the consumer or enterprise market trends. But overall, industrial sectors do follow the general trends in communication technologies. Key trends and issues include followings

- Continued use of fieldbus technology: The fieldbus technologies are the digital LAB of the industry. They connect the sensors, controllers, and actuators of most factory automation and process control facilities. Despite the ongoing movement to Ethernet connectivity and wireless, there continues to be the growth of several percents per year in the fieldbus market.
- A strong movement to Ethernet: Ethernet has been the local area network (LAN) of choice for enterprise and even consumer networking for decades, and it dominates. The industry was slow to adopt it but has now embraced it completely. Most new

industrial networking efforts use some form of Ethernet. Its proven reliability, low cost, and high availability have made it particularly popular. Special industrial versions of Ethernet have emerged to enhance it for industrial use.

- Significant growth in wireless connectivity: Industry was slow to adopt wireless despite its many benefits. Industrial users assumed it was unreliable and insecure but have learned otherwise since. New and improved wireless standards and equipment have made wireless a key component in most modern industrial settings.
- Fewer proprietary standards and equipment: For decades, industrial communications needs were met with many high-cost proprietary fieldbuses, interfaces, and equipment, which are still entrenched in many systems. However, the trend today is to open standards and Ethernet.
- Rapid adoption of the Internet protocol (IP) model: The goal is to give the most industrial equipment an IP address so devices and equipment can communicate over Ethernet and the Internet. With the availability of IPv6, that is now possible.
- Increased use of video surveillance: security has become an issue at many plants and facilities, and video is useful. Video also enables improved monitoring that simple sensors cannot provide.
- Industrial Standardization for Interoperability: Most factories, process control plants, and facilities are a real mixed bag of old and new, analog and digital, and proprietary and open standards. A big issue has been the incompatibility and interoperability of different equipment such as all devices and system can work together seamlessly. Such challenges lead to new standards, equipment, and software gradually developed to address those problems.

1.6 Conclusions

Technology is continuously and rapidly transforming industrial processes. It is sometimes hard for businesses to integrate a new technology into an existing system. It requires professional expertise and training to run a newly introduced system. However, with the growing demand for sophisticated and high-quality products, businesses have to quickly adapt and utilize the power of emerging automation systems.

Looking to the future, the most notable trend appearing in the industry is the move to industrial wireless networks at all levels [8–10]. Wireless networks further reduce the volume of wiring needed (although oftentimes power is still required), enable the placement of sensors in difficult locations, and better enable the placement of sensors on moving parts such as on tooltips that rotate at several thousand revolutions per minute. Issues with the migration to wireless include interference between multiple wireless networks, security, and reliability and determinism of data transmission. The anticipated benefit in a number of domains (including many outside of manufacturing) is driving innovation that manufacturing, in general, can leverage. It is not inconceivable that wireless will make significant in-roads into networked control and even safety over the next 5–10 years

References

1. Voss W The future of CAN/CANopen and the industrial ethernet challenge. ESD Electronics, Inc., USA. http://www.rtcgroup.com/whitepapers/files/TheFutureofCAN.pdf
2. Brooks P (2001) Ethernet/IP: industrial protocol. Retrieved 10 Jan 2005, from http://literature.rockwellautomation.com/idc/groups/literature/documents/wp/enet-wp001_-en-p.pdf
3. Bettendorf B (2004) Which industrial network for you (Industrial Networking). Putman Media Publications, p 34
4. Stenerson J (1999) Industrial networks. In: Stewart CE (ed) Sensors and communications fundamentals of programmable logic controllers (2nd edn). Prentice Hall, Upper Saddle River, New Jersey, pp 367–392
5. Kopetz H (1991) Event-triggered versus time-triggered real-time systems. In: Lecture notes in computer science, operating systems of the 90s and beyond, vol 563. Springer, Heidelberg, Germany, pp 87–101
6. Berge J (2001) Fieldbuses for process control: engineering, operation, and maintenance. ISA: Research Triangle Park, NC, USA
7. Open DeviceNet Vendors Association (ODVA). https://www.odva.org/
8. Galloway B, Hancke GP (2013) Introduction to industrial control networks. IEEE Commun Surv Tutorials 15(2):860–880
9. Willig A, Matheus K, Wolisz A (2005) Wireless technology in industrial networks. Proc IEEE 93(6):1130–1151
10. Djiev S. Industrial networks for communication and control. https://data.kemt.fei.tuke.sk/SK_rozhrania/en/industrial%20networks.pdf

Chapter 2
FlexRay Protocol: Objectives and Features

2.1 Introduction

In industrial networks, many mechanical controlling parts have been replaced with electronic control units (ECU) . Many ECUs are implemented in industrial automation modes, and the number is still increasing. The communication networks in vehicles transmit signal data that are encapsulated in messages. Most of these messages are real-time messages, i.e., their timely delivery must be guaranteed. Technically, precomputed message schedules have to be supplied to meet such timing requirements [1]. In addition, considering the fast growth in the number of ECUs and signals in automotive electronics, the communication must be efficient to provide system extensibility. As a result, the complexity of industrial networks is increasing rapidly. Different types of communication protocols are currently being used in different automobiles such as Controller Area Network (CAN) and FlexRay.

FlexRay is a new high-bandwidth communication protocol for the automotive domain. It is expected as the next generation bus for automotive industry and to be the de facto standard for high-speed. Most remarkable features of FlexRay can be related such as high data rate, time/event-triggered behavior, deterministic, fault-tolerance, and redundancy. Along with the development of the automotive industry for higher requirements as safer, more comfortable, reliable, complex demands at same time and enhancing memory, the development of electronic control unit (ECU) also plays an important role. More and more ECUs have been used widely nowadays and it has been developed from 8 bits up to 32 bits.

FlexRay protocol can implement both time-triggered and event-triggered message. In the first case, task activations and frame transmissions are bound to happen at predefined points in time. In the other case, the sporadic real-time messages are generated by event occurrences and have to be transmitted before their deadline. According to these methods, FlexRay can be divided to two main fields of priority assignment to jobs: static segment (SS) and dynamic segment (DS). The organization of SS is based on a Time-Division Multiple Access (TDMA) operation. It consists

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_2

of a fixed number of equal-size static slots (STSs) and transmits message following uniquely frame identifiers (FIDs). In static method, the fixed priority is selected for each job at the beginning of the development. Besides, the DS uses the flexible TDMA (FTDMA) scheme. In contrast to SS, the time when the message begins to transmit is not fixed, but the priority of the message is fixed with DS [2]. DS is conducted in dynamic slots (DYSs) that are superimposed on mini slot (MS). When message is transmitted by DYS, the length of the DYS is equal to the number of MS for message transmission. Otherwise, the length of the DYS is one MS.

2.2 FlexRay System

FlexRay was developed for next-generation automobiles by a consortium founded by BMW, Bosch, DaimlerChrysler and Philips in 2000. FlexRay is a new standard of network communication system which provides a high-speed serial communication, time-triggered bus, and fault-tolerant communication between electronic devices for future automotive applications. FlexRay supports a time-triggered scheme and an optional event-triggered scheme. The upper bound of the data rate is 10 Mbps and it provides two channels for redundancy (FlexRay Consortium, 2005). FlexRay protocols are first designed using SDL (Specification and Description Language). Then, the system is re-designed using Verilog HDL based on the SDL source. In addition, FlexRay system is combined with active stars. The combined system is implemented using ALTERA Excalibur ARM EPXA4F672C3. It is shown that the implemented system operates successfully. FlexRay architecture is divided into three levels in Fig. 2.1 based on protocol [3].

2.2.1 Level 1—Network Topology

In this section, level 1 introduces the network topology overview for FlexRay. It supports single/dual channel and three types of topology: bus-type, star-type, and hybrid-type combined with bus-type and star-type. The dual channel is fault tolerant by making redundant configuration. In addition, the star-type topology is used to connect by the point-to-point for high-speed data rate, and it is easy to reduce failures.

2.2.2 Level 2—Interface

In level 2, FlexRay interface supports bus guardian at physical interface including error containment and error detection in the time domain. Then, the bus guardian interacts with both communication controller and host processor.

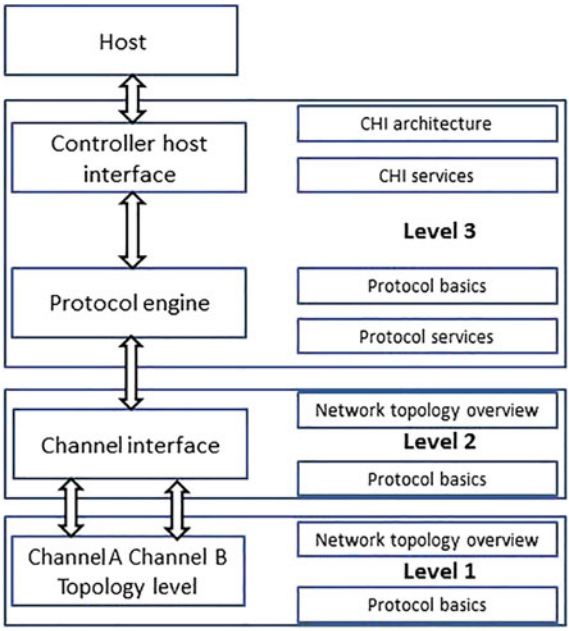


Fig. 2.1 FlexRay architecture levels

2.2.3 Level 3—CHI and Protocol Engine

Level 3 contains the control host interface (CHI) and protocol engine. The protocol engine is also called communication controller (CC). The node architecture is shown in Fig. 2.1. According to [3], each node consists of a host and a communication controller (CC) which is connected by a controller–host interface (CHI). The CHI serves as a buffer between the host and the CC. The host is the part of ECD where processes incoming messages and generates outgoing messages. The CC independently implements FlexRay protocol services.

In FlexRay protocol, media access control is based on a communication cycle as Fig. 2.2. The cycle comprises a static segment (SS), a dynamic segment (DS), a symbol window (SW), and the network idle time (NIT). The organization of the SS is based on a time-division multiple access (TDMA) scheme. It transmits message according to the fixed number of equal-size static slots (STS) and uniquely frame identifiers (FIDs). The DS employs the flexible TDMA (FTDMA) approach. It is divided into mini slots. The communication in the dynamic segment is conducted in dynamic slots (as opposed to static slots of fixed size in the static segment). The SW and the NIT provide time for the transmission of internal control information and protocol-related computations.

As can be seen in Fig. 2.3, FlexRay frame includes three segments: header, payload, and trailer segment. The first five bits are defined as the basic of the frame. For static segment, Frame ID (11 bits) is defined as the slot position. In dynamic

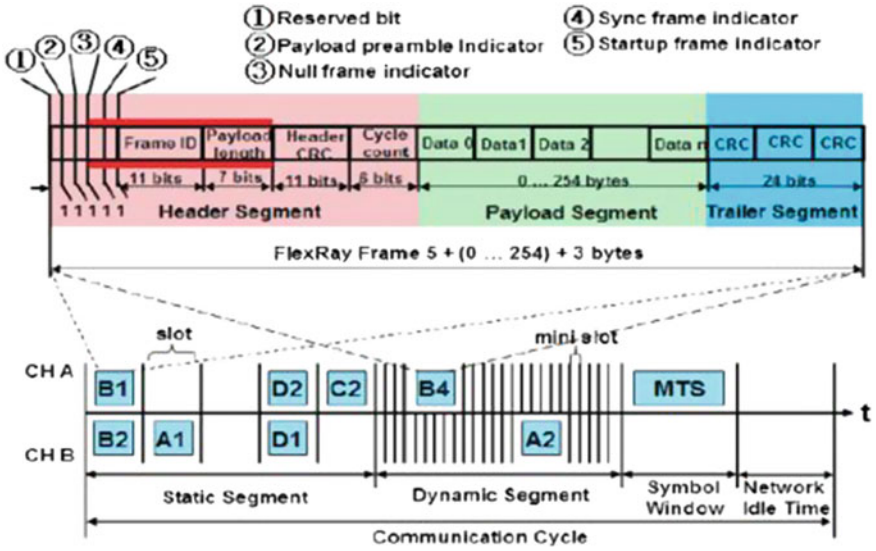


Fig. 2.2 FlexRay frame format and communication cycle

segment, Frame ID is used to indicate the priority of the frame: a lower identifier indicates higher priority.

In header segment, payload length (7 bits) is the data length (payload length × 2 = number of data bytes). Header CRC (11 bits) is defined as the Cyclic Redundancy Check, which is computed over the sync frame indicator (1 bit), startup frame indicator (1 bit), frame ID (11 bits) and payload length (7 bits). Cycle count (6 bits) is the serial number of the frame that defines locally in the node. Payload segment (0–254 bytes) contains the main data which is transferred via bus. Trailer segment (24 bits) is used for cyclic redundancy check, it is computed over the header segment and payload segment.

2.3 Message Scheduling for FlexRay System

2.3.1 FlexRay Static Segment

There are several methods to optimize scheduling for FlexRay in which each algorithm is presented relying on the specify model, hence the special method is directed toward the different model. Pop et al. have introduced the scheduling model that divides into three steps. First, the mapping step including time-triggered cluster and event-triggered cluster, the second step is frame packing, and the final step is time scheduling. For each step, a given set of parameters is leading to find out if a sys-

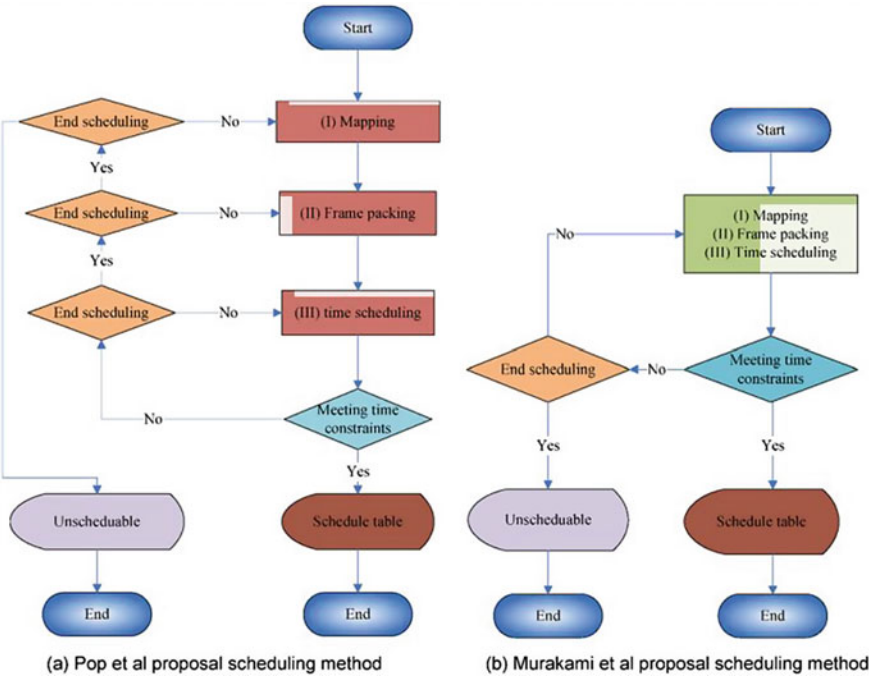


Fig. 2.3 Comparison of the scheduling models

tem is schedulable and the constraints are met. If it is unscheduled in any steps, the set of parameters must be changed until obtaining the approximate optimal solution [4]. As Fig. 2.3 shown, Pop et al. divide the scheduling model into three steps: the mapping including time-triggered cluster and event-triggered cluster, frame packing, and time scheduling. For each step, a given set of parameters is leading to find out if a system is schedulable, and the constraints are met. If it is unscheduled in one of these steps, the set of parameters must be changed until obtaining the approximate optimal solution.

Besides, Murakami et al. also propose a simulated annealing (SA) algorithm as a static scheduling method for FlexRay system [5]. This method is based on the scheduling model of Pop et al. but combining three steps: Mapping, Frame packing, and Time scheduling into only one step. The scheduling method of Murakami et al. suppose that all the processes belong a process graph can have difference periods and also take into account the influence of system load on the scheduling method [6]. Then, the genetic algorithm (GA) for FlexRay is presented and optimized the formal model called the data flow diagram (DFG). The GA is better than SA algorithm approach. First, authors give a brief description of the problem based on TDMA for FlexRay static segment and consider the synchronous input/output constraints. Then, based on the systematic scheduling model, they use the GA to optimize the

tasks assignment in ECUs and message scheduling through bus. After that, GA and hybrid-GA are proposed in [6].

In these works above, none of the approaches accounts for the potential jitter in the message transmission. The message schedule computation for the SS of FlexRay is designed to accommodate periodic real-time messages. Previous work on this topic focuses on the timing analysis of applications on a FlexRay bus or on heuristic strategies that aim at finding a feasible message schedule for a given message set. Klaus Schmidt and Ece Guran Schmidt identify and solve two issues of message scheduling on SS of FlexRay in [7]. The signals have to be packed into equal-size messages to obey the restrictions of FlexRay protocol, while using as little bandwidth as possible and message schedule has to be determined such that the periodic messages are transmitted with minimum jitter. To solve these issues, the authors formulate a nonlinear integer programming (NIP) problem to maximize bandwidth utilization. Besides, they also introduce appropriate software architecture and derive an integer linear programming (ILP) problem that both minimize the jitter and the bandwidth allocation.

In addition, to minimize the number of used slots in the mentioned above issue, the authors also present a formulation for the minimization of the transmission jitter. This works discuss optimization of the static segment communication schedule, but do not attempt optimization at the system-level. They do not consider possible end-to-end deadlines, information passing and precedence constraints among tasks and signals, nor synchronization of the task and signal schedules. To solve these problems, Haibo Zeng et al. study the problem of the ECU and FlexRay bus scheduling synthesis from the perspective of the application designer, interested in optimizing the scheduling subject to timing constraints with respect to latency or extensibility-related metric functions [8]. They introduce solutions for a task and signal scheduling problem, including different task scheduling policies based on existing industry standards. The solutions are based on the Mixed-Integer Linear Programming (MILP) optimization framework instead of a heuristic method to schedule transactions consisting of tasks and signals on a FlexRay-based system. This formulation includes the system-level schedule optimization with the definition of an optimal relative phase in the activation of tasks and signals which accounts for deadlines and precedence constraints. The objective of the proposed MILP method is to maximize the number of free communication slots and improve extensibility or to maximize minimum laxity among paths and improve timing performance. The authors provide solutions for the synchronized task-to-signal information passing under different task scheduling policies based on existing industry standards. There are many algorithms to optimize scheduling for FlexRay based on the different models. It is not only the different optimization algorithms, but also different perspectives including models and methods.

2.3.2 FlexRay Dynamic Segment

The dynamic segment of FlexRay is based on even-triggered and the message is acyclic and arrives at any time. Therefore, a probabilistic model is needed to analyze message transmission in the dynamic. This section is related research on schedulability analysis and optimization algorithms for FlexRay dynamic segment. Nielsen et al. propose a novel approach to performance analysis of the dynamic segment based on Markov chain transient analysis [9]. The model of the dynamic segment is a two-dimensional discrete-time Markov chain, where the discrete time steps represent mini slots of the dynamic segment. The state space of this model is composed of slot identity and states within a dynamic slot, each state in the Markov chain constitutes either an idle mini slot, or a mini slot that is used for transmitting a frame. The time step of this model has the duration of one mini slot. The transition probabilities of the Markov chain correspond to the arrival probabilities of each dynamic slot. Then the transition probability matrix can be calculated in the way of iteration over the set of dynamic slots, and the state probability vector should be calculated easily. The distribution of last dynamic slot can be taken from the Markov chain model. Hence, the model based on Markov chain can be a tool for network designer to make early predictions on network behavior.

According to Nielsen et al. model, the authors in [10] improve the Markov chain model, and propose the probabilistic delay model of dynamic segment message. The delay model considers variable length messages that share same Frame ID. While the delay model assumes that frames within a frame ID use the same payload length in [9], in this innovated model, the messages share a frame ID has the same choice to be sent on the bus. The authors focused on the bus utilization and the delay time in different angles, they presented two performance metrics as analysis targets: frame delay probability and empty mini slot distribution.

The dynamic messages produce and transmit between two tasks each message has a different length. For given FlexRay system, the length of dynamic segment is fixed, while during the dynamic segment, if no message is to be sent during a certain slot, then that slot will have a length of one mini slot, otherwise the dynamic slot will have a length equal with the length of message transmitted. Based on the characteristic of the dynamic segment, many researchers describe the optimization of target with bus bandwidth, response time and distribution of slot. This chapter just introduces the work recently on bus bandwidth.

In term of bus bandwidth, Ece Guran Schmidt and Klaus Schmidt consider the bounds on the message generation times and the timing requirements for message delivery of the sporadic messages to reserve bandwidth for each message based on addressing a message schedule method for the sporadic message [11]. In the authors' point of view, the DS of FlexRay is designed to accommodate sporadic real-time messages that are generated by event occurrences and have to be transmitted before their deadline. To this end, it is required to find feasible message schedules that meet the timing requirements. Previous work on FlexRay DS mostly provides methods to test if a given schedule is feasible. They propose a method for synthesizing

efficient and feasible message schedules. Based on a formal problem description, their approach determines the required system parameters such that the sporadic messages are delivered on time. Two performance metrics are defined to measure the efficiency of each schedule: the bandwidth reservation and the cycle load. The bandwidth reservation indicate the number of mini slots reserved per cycle for each node, while the cycle load denotes the maximum number of mini slots that is reserved for message transmission in a cycle. Then, integer programming is applied to group the messages, and the messages in same group will be transmitted in a reservation. The optimal schedule satisfies more reservations utilized by the groups of message and bandwidth reservation is minimized.

2.3.3 Comparison with CAN

2.3.3.1 Physical Layer

As mentioned before, FlexRay supports three main topologies, bus, star and hybrid and its baud rate is 10 Mbps while CAN only use bus as its topology with 1 Mbps baud rate. CAN only uses one channel and using metal as its physical layer while FlexRay can support two channels on its implementation on metal or optical fiber. CAN bus length can reach 40 m, and FlexRay only 22 m between each node, or node and active-star.

2.3.3.2 Communication

FlexRay communication is based on time-triggered for periodic message and event triggered for aperiodic message. FlexRay can connect up to 22 nodes in bus or star topology. CAN only use event triggered for its communication and can connect to nodes depending on delay time of the bus.

2.3.3.3 Frame

FlexRay frame has three segments. Its maximum payload is 254 bytes, while CAN only has 8 bytes of payload. CAN frame consists of Data frame, Remote frame, Error Frame, and Over Rode.

2.3.3.4 Error Condition

CAN has five types of error bit error, stuffing error, ACK error, framing error, and CRC error. FlexRay has all type of errors except clock synchronization. In CAN, error status divided in three steps, error active, error passive, and bus off. FlexRay error

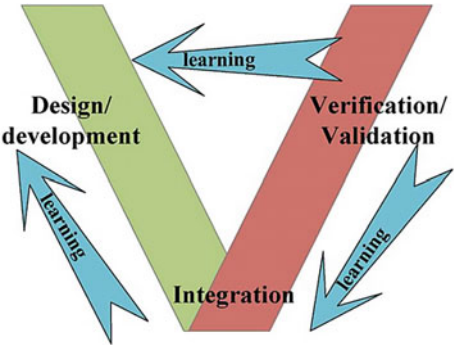


Fig. 2.4 Modem V diagram

status also divided into three steps, normal active, normal passive, and halt. Active condition in both FlexRay and CAN means there are no errors. Passive condition means that errors in node are still acceptable. Bus off or halt condition mean the error occurred is fatal.

2.4 Verification and Validation

FlexRay network will be extensively used in safety and security critical areas such as powertrain management system. An error in any of these systems can cause not only huge financial loss, but also loss of human lives. The modern V diagram model help engineer to develop FlexRay network quickly as Fig. 2.4.

In Fig. 2.4, the design/development stage (top left) and verification and validation stage (top right) are implemented at the same. Both of them are gradually to meet the demand of the application via the integration. The modem V diagram corresponds to engineering processes is iterative with many repetitive steps throughout the development life cycle. At last, the development life cycle completes the entire design. Then the verification and validation is important for FlexRay. There are some validation methods including computer simulation, formal verification.

2.4.1 Computer Simulation for Model Validation

Computer simulation for model verification includes scheduling analysis (on abstract target model), and time simulation and analysis (on detailed target model). Usually, scheduling analysis that abstracts model is in the early design stage and time simulation and analysis is in the late design stage for certainty model.

2.4.1.1 Scheduling Analysis for Abstract Target Model

In the early design stage, model-based design is universally recognized to be a good approach to schedule synthesis and increasingly used in the automotive industry. For instance, the data flow diagram modeling abstracts and defines the system's necessary functions. In all mentioned contributions for static and dynamic segment, the basic idea is to create a formal model that covers the relevant load arrival into the system. Models of the plant interacting with the scheduling algorithm are available in the early design. In that mean, only some adjustment of sensitive parameters could be sufficient to obtain reliable models. It clearly focuses on timing-relevant influences resulted in very small and efficient models with only few parameters. Such as slot size, frame cycle times in FlexRay [12]. The key parameters from powerful abstraction described the arrival of system load without details of the executed code and the content of transmitted data. It is enable for that the systematic and efficient analysis of timing and performance, and the early application depending on V diagram for next stage analysis and design.

2.4.1.2 Time Simulation and Analysis for Certainty Model

Time simulation and analysis, in FlexRay, work at the level of bus protocol and operating system, and analyses timing effects resulting from integration of tasks on ECU and messages on buses, respectively. Based on the executable system model, scheduling analysis is applicable in different level and tells about performance reserves or upcoming bottlenecks before the entire system is build.

2.4.2 Formal Verification

Including both the hardware and software systems, formal verification is the action to prove or disprove the correctness of algorithms, it is done by providing a formal proof on an abstract mathematical model of the system. Then there are many works on formal verification of hardware, software, and protocols for FlexRay. It becomes even more critical when considering FlexRay distributed embedded systems. Erik Enders etc. consider the close interaction of software and hardware parts, and verify against the specification as seen by a system programmer [13]. The formal verification of FlexRay protocol was introduced [14]. It verifies the bus guardian properties and the clock synchronization algorithm. Rush by gives an overview of the formal verification of a Time-Triggered architecture and formally proves the correctness of some key algorithms. The goal of formal verification is feasible to formally verify a complex distributed automotive system in a pervasive manner. The desired goal is a "single" top-level theorem that describes the correctness of the whole system.

2.5 Software and Hardware

FlexRay technology can be split into two (software/hardware) to main areas: software to configure and manage both communication in a FlexRay cluster and the application layer of ECU; digital and logic implementing FlexRay protocol and analog signal drivers.

2.5.1 Software

The software used in today's automobile is highly heterogeneous. The result is that different developers use different standards to develop same software and different suppliers develop software components for different hardware. It includes different types of electronic control units as well as different types of buses. Therefore, the automobile manufacturers, suppliers, and tool developers jointly have been develop open and standardized automotive software architecture (AUTOSAR). The objective of the AUTOSAR initiative is establishing an open standard for automotive electric/electronic architectures. The scope of AUTOSAR includes all vehicle domains as following:

1. Implementation and standardization of basic system functionalities as an OEM wide "Standard Core" solution.
2. Scalability to different vehicle and platform variants.
3. Transferability of functionalities throughout network.
4. Integration of functional modules from multiple suppliers.
5. Consideration of availability and safety requirements.
6. Redundancy activation.
7. Maintain ability throughout the whole "Product Life Cycle".
8. Increased use of "Commercial off the shelf hardware".
9. Software updates and upgrades over vehicle lifetime.

Base on Fig. 2.5, AUTOSAR standard will serve on different hardware platform in the future vehicle applications and serve to minimize the current barriers between functional domains. It is also to map functional networks to different ECUs in the system. For the technical goal modularity, scalability, transferability and re-usability of functionalities, AUTOSAR provide common software infrastructure for automotive systems of all vehicle domains based on standardized interfaces [15].

2.5.2 Hardware

The center of the hardware for FlexRay is the protocol execution layer, where outgoing frame data is sent to the physical layer. The physical layer contains three parts:

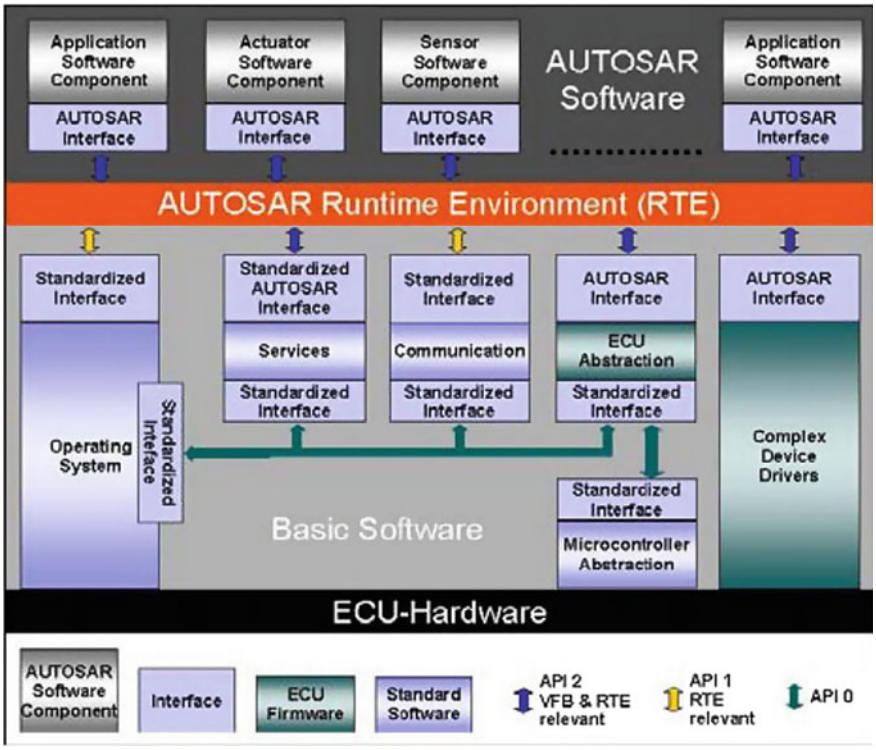


Fig. 2.5 AUTOSAR ECU software architecture

the bus drivers, the optional bus guardians, and the physical interconnections. The interfaces for integrating FlexRay controller into the system as follows:

- Clock and Reset Interface: enables clock gating and reset control through hard or soft resets.
- Host Interface: a simple read/write peripheral interface.
- Interrupt and Strobes Interface: selects interrupt and debugging implementations through software.
- FlexRay Bus Interface: used to connect FlexRay device to FlexRay bus drivers, specified in FlexRay Communication System Electrical Physical Layer Specification.
- System Memory Interface: connected through the Bus Master Interface (BMIF) to an external memory controller. This can be connected directly to a shared memory or an external memory bus subsystem. In either case, certain latency requirements must be met.

Many semiconductor vendors are working on FlexRay controller development and integrate FlexRay controller with MCU. To develop FlexRay-based system, Freescale, Fujitsu, Atmel, etc., offer multifunctional evaluation boards with the 16/32

bit MCU as a host controller. Infineon and NXP respectively have developed CC for FlexRay bus communication. In the 32-bit MCU domain, Freescale develops MPC5567 that is based on the PowerPC core and enables fault-tolerant communication at high-bandwidth rates of 10 Mbps, reducing system cost by integrating maximum functionality on the chip. This device is the first 32bits flash-based MCU with FlexRay protocol. Its integrated FlexRay functionality is that control modules, deterministic and dependable manner based on FlexRay protocol in the car. This helps to popularize FlexRay application in braking, stability, and suspension system. FlexRay is gaining international support within the automotive industry and will be used by vehicle makers to enable new safety-critical and performance features.

2.6 Conclusions

This chapter presents a situation from the scheduling analysis at the early design stage to the scheduling/optimization algorithms for the static and dynamic segment. There are many works focus on scheduling analysis for FlexRay system. Each approach to optimize scheduling method relies on specific model systems or applications. This chapter presented briefly the situation of study from the scheduling analysis at the early design stage to the scheduling/optimization algorithms for the static and dynamic segment. In the future, when FlexRay becomes as data backbone, the work on FlexRay not only analyze the service condition of individual bus, but also research the entire network design including different topologies of CAN, LIN and FlexRay bus to develop the gate routing protocol in order to connecting other buses via gateway.

References

1. Park H-S, Kim D-S, Kwon W-H (2002) A scheduling method for network-based control systems. IEEE Trans Control Syst Technol 10(3):318–330. IF: 2.521, ISSN: 1063–6536
2. He X, Wang Q, Zhang Z (2011) A survey of study of Flexray systems for automotive net. In: International conference on electronic and mechanical engineering and information technology (EMEIT), vol 3, pp 1197–1204
3. Temple C (2003) Protocol overview. In: FlexRay international workshop, Detroit
4. Pop T, Pop P, Eles P, Peng Z, Andrei A (2006) Timing analysis of Flexray communication protocol. In: 18th euromicro conference on real-time systems, pp 11, 216
5. Murakami HTMK, Liyama S, Hosotani I (2007) A static scheduling method for distributed automotive control systems. IPSJ Trans Adv Comput Syst 48:203–215
6. Ding S, Tomiyama H, Takada H (2008) An effective ga-based scheduling algorithm for Flexray systems. IEICE Trans Inf Syst E91-D(8):2115–2123
7. Schmidt K, Schmidt E (2009) Message scheduling for Flexray protocol: the static segment. IEEE Trans Veh Technol 58(5):2170–2179
8. Zeng H, Di Natale M, Ghosal A, Sangiovanni-Vincentelli A (2011) Schedule optimization of time-triggered systems communicating over Flexray static segment. IEEE Trans Industr Inf 7(1):1–17

9. Jessen J, Peter Schwefel NH, Hamdan A (2007) Markov chain-based performance evaluation of Flexray dynamic segment. In: 6th international workshop on real time networks (RTN 07)
10. Kim B, Park K (2009) Probabilistic delay model of dynamic message frame in Flexray protocol. IEEE Trans Consum Electron 55(1):77–82
11. Schmidt E, Schmidt K (2009) Message scheduling for Flexray protocol: the dynamic segment. IEEE Trans Veh Technol 58(5):2160–2169
12. Ding S, Murakami N, Tomiyama H, Takada H (2005) A GA-based scheduling method for FlexRay systems. In: Proceedings of the 5th ACM international conference on embedded software, ser. EMSOFT '05, ACM, New York, USA, pp 110–113
13. Endres E, Mller C, Shadrin A, Tverdyshev S (2010) Towards the formal verification of a distributed real-time automotive system. In: NASA formal methods, ser. NASA conference proceedings, NASA/CP-2010-216215, pp 212–216
14. Zhang B (2006) On the formal verification of flexray communication protocol. In: Merz S, Nipkow T (eds) Automatic verification of critical systems, Nancy/France, pp 184–189
15. Bunzel S (2011) Autosar—the standardized software architecture. Informatik-Spektrum 34(1):79–83

Chapter 3
Communication Using Controller Area Network Protocol

3.1 Introduction

Controller Area Network (CAN) was originally developed in February 1986, by Robert Bosch Gmbh and was introduced as "Automotive Serial Controller Area Network" at the SAE congress in Detroit as the new bus system. At the beginning of the 1980s, a group of engineers at Bosch GmbH were the pioneers in introducing this multi-master network protocol. It was based on a nondestructive arbitration mechanism that grants bus access to the message without causing any delays. This Automotive Serial CAN protocol was introduced due to the fact that the cars required the extra wiring costs for increased number of distributed control system and also that none of the existent protocols could perform satisfactorily for the automotive engineers [1].

At the beginning, CAN was used in interconnecting the ABS (Anti-Block System and Acceleration Skid Controls (ASC). For example in ASC, engine timing and carburetor control are required when slippage occurs and vice versa. It was first developed for the automotive industry, other automation sectors and today is used in the other large variety of embedded systems applications [2, 3]. The first hardware implementation of CAN protocol was produced by Intel Corporation in mid-1987 in the form of controller chip, the 82,526 which favored the FullCan concept as compared to BasicCAN implementation introduced by Phillips Semiconductors which shortly followed. The semiconductor vendors who implemented CAN modules into their devices were mainly focused on the automotive industry and since the mid-1990s, Infineon Technologies (formerly Siemens Semiconductors) and Motorola have manufactured and delivered large quantities of CAN controller chips to the European passenger car manufacturers and their suppliers. Even though it was conceived for vehicle applications, at the beginning of the 1990s, CAN began to be adopted in different scenarios. The standard documents provided satisfactory specifications for the lower communication layers but did not offer guidelines or recommendations for the upper part of the Open Systems Interconnection (OSI) protocol stack, in general,

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_3

and for the application layer, in particular. This is why the earlier applications of CAN outside the automotive scenario (i.e., textile machines, medical systems, and so on) adopted ad hoc monolithic solutions.

CAN protocol can be interpreted as being a two layer protocol in terms of the OSI/ISO 7-layer reference model. In other words, CAN operates at the physical layer and the data link layer of the standard OSI model. The physical layer determines how the signal is transmitted. The International Standards Organization (ISO) defined a standard ISO11898 which incorporates the CAN specifications to meet some of the requirements in the physical signaling [4], which includes bit encoding and decoding (Non-Return-to-Zero, NRZ) as well as bit timing and synchronization [5]. Using serial bus network mechanisms, the existing CAN applications send messages over the network [6]. In the CAN systems, there is no need for central controller as every node is connected to the every other node in the network. CAN communications protocol, ISO 11898: 2003, describes how information is passed between devices and conforms to the Open Systems Interconnection (OSI) model that is defined in terms of layers. Actual communication between devices by the physical medium is defined by the physical layer of the model. CAN specifications [7, 8], in particular, include only the physical and data link layer as shown in Fig. 3.1.

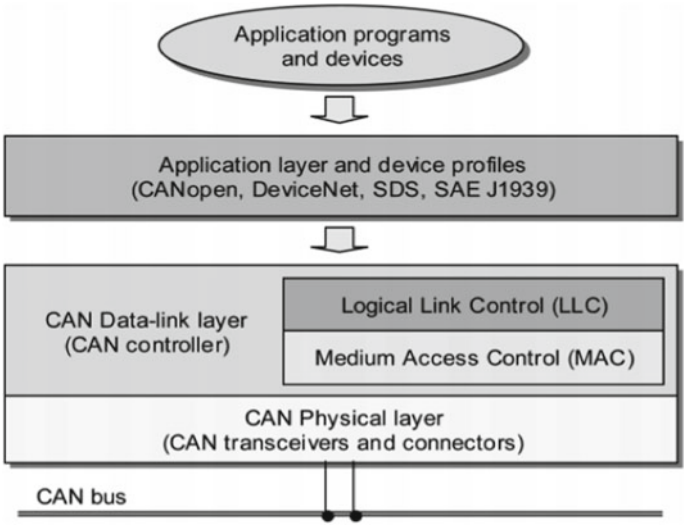


Fig. 3.1 CAN protocol stack

3.2 CAN Protocol Overview

3.2.1 Physical Layer

The features of the physical layer of CAN that are valid for any system, such as those related to the physical signaling, are described in ISO 11898-1 [7]. The medium access units (i.e., the transceivers) are defined in two separate documents: ISO 11898-2 [8] and ISO 11898-3 [9] for high-speed and low-speed communications, respectively. The definition of the medium interface (i.e., the connectors) is usually covered in other documents.

3.2.1.1 Network Topology

CAN networks are based on a shared-bus topology. Buses have to be terminated at each end with resistors (the recommended nominal impedance is 120 W), so as to suppress signal reflections. For the same reason, the standard documents state that the topology of a CAN network should be as close as possible to a single line. Stubs are permitted for connecting devices to the bus, but their length should be as short as possible. For example, at 1 Mbit/s the length of a stub must be shorter than 30 cm.

Several bit rates are available for the network, the most adopted being in the range of 50 Kbit/s to 1 Mbit/s (the latter value represents the maximum allowable bit rate according to the CAN specifications). The maximum extension of a CAN network depends directly on the bit rate. The exact relation between these two quantities involves parameters such as the delays introduced by transceivers and optocouplers. In general, the mathematical product between the length of the bus and the bit rate has to be approximately constant. For example, the maximum extension allowed for a 500 Kbit/s network is about 100 m, and increases up to about 500 m when a bit rate of 125 Kbit/s is considered.

Signal repeaters can be used to increase the network extension, especially when large plants have to be covered and the bit rate is low or medium. However, they introduce additional delays on the communication paths; hence the maximum distance between any two nodes is effectively shortened at high bit rates. Using repeaters also achieves topologies different from the bus (trees or combs, for example). In this case, good design could increase the effective area that is covered by the network.

3.2.1.2 Bit Encoding

In CAN, the electrical interface of a node to the bus is based on an open-collector-like scheme. As a consequence, the level on the bus can assume two complementary values, which are denoted symbolically as dominant and recessive. Usually, the dominant level corresponds to the logical value 0 while the recessive level coincides with the logical value 1.

CAN relies on the non-return-to-zero (NRZ) bit encoding, which features very high efficiency in that synchronization information is not encoded separately from data. Bit synchronization in each node is achieved by means of a digital phase-locked loop (DPLL), which extracts the timing information directly from the bit stream received from the bus. In particular, the edges of the signal are used for synchronizing the local clocks, so as to compensate tolerances and drifts of the oscillators.

3.2.2 Message Frame Format

CAN protocol supports two message frame formats, the main difference is the length of the identifier field and some other bits in the arbitration field. In particular, the standard frame format (also known as CAN 2.0A format) defines an 11-bit identifier field, which means that up to 2048 different identifiers are available to the applications executing in the same network (many older CAN controllers only support identifiers in the range of 0–2031). The extended frame format (identified as CAN 2.0B) instead assigns 29 bits to the identifier, so that up to a half billion different objects could exist (in theory) in the same network. This is a fairly high value, which is virtually sufficient for any kind of application.

3.2.2.1 Data Frame

Each data frame in CAN begins with a start-of-frame (SOF) bit at the dominant level, as shown in Fig. 3.2. Immediately after the SOF bit there is the arbitration field, which includes both the identifier and the remote transmission request (RTR) bit. As the name suggests, the identifier field identifies the content of the frame that is being exchanged uniquely on the whole network. The identifier is also used by the MAC sub-layer to detect and manage the priority of the frame, which is used whenever a collision occurs (the lower the numerical value of the identifier, the higher the priority of the frame).

The identifier is sent starting from the most significant bit up to the least significant one. The size of the identifier is different for the standard and extended frames. In the latter case, the identifier has been split into an 11-bit base identifier and an 18-bit extended identifier, to provide compatibility with the standard frame format.

The RTR bit is used to discriminate between data and remote frames. Since a dominant value of RTR denotes a data frame while a recessive value stands for a remote frame, a data frame has a higher priority than a remote frame having the same identifier.

Next to the arbitration field comes the control field. In the case of standard frames, it includes the *identifier extension* (IDE) bit, which discriminates between standard and extended frames, followed by the reserved bit r0. In the extended frames, the IDE bit effectively belongs to the arbitration field, as well as the *substitute remote request*

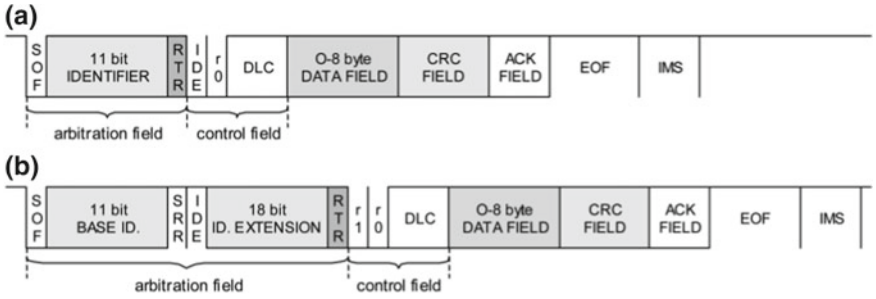


Fig. 3.2 Format of data frames

(SRR) bit—a placeholder that is sent at recessive value to preserve the structure of the frames. In this case, the IDE bit is followed by the identifier extension and then by the control field, which begins with the two reserved bits r1 and r0. After the reserved bits there is *the data length code* (DLC), which specifies—encoded on 4 bits—the length (in bytes) of the data field. Since the IDE bit is dominant in the standard frames, while it is recessive in the extended ones, when the same base identifier is considered, standard frames have precedence over extended frames.

After the data field there are the CRC and acknowledgment fields. The former field is made up of a cyclic redundancy check sequence encoded on 15 bits, which is followed by a CRC delimiter at the recessive value. The kind of CRC adopted in CAN is particularly suitable to cover short frames (i.e., counting less than 127 bits). The acknowledgment field is made up of two bits: the ACK slot followed by the ACK delimiter. Both of them are sent at the recessive level by the transmitter. The ACK slot, however, is overwritten with a dominant value by each node that has received the frame correctly (i.e., no error was detected up to the ACK field). It is worth noting that ACK slot is actually surrounded by two bits at the recessive level: CRC and ACK delimiters. By means of ACK bit, the transmitting node is enabled to discover whether at least one node in the network has received its frame correctly.

At the end of the frame, there is the end-of-frame (EOF) field, made up of seven recessive bits, which notifies all the nodes of the end of an error-free transmission. In particular, the transmitting node assumes that the frame has been exchanged correctly if no error is detected until the last bit of the EOF field, while in the case of receivers, the frame is valid if there are no errors until the sixth bit of EOF.

3.2.2.2 Remote Frame

The main duty of remote frame is to solicit the transmission of data from another node. On the one hand, this type of message is explicitly marked as a remote frame by a recessive RTR bit in the arbitration field. Remote frames are used to request that a given message be sent on the network by a remote node. It is worth noting that the requesting node does not know who the producer of the related information is.

It depends on the receivers to discover the one that has to reply. The DLC field in remote frames is not effectively used by the CAN protocol. However, it should be set to the same value as the corresponding data frame, so as to cope with the situations where several nodes send remote requests with the same identifier at the same time. In this case, it is necessary for the different requests to be perfectly identical, so that they will overlap in the case of a collision.

3.2.2.3 Error Frame

Error frame is a special message that violates the formatting rules of a CAN message. It is transmitted when a node detects an error in a message, and causes all other nodes in the network to send an error frame as well. The original transmitter then automatically retransmits the message. Error frames consist of two fields: error flag and error delimiter. There are two kinds of error flag: the active error flag is made up of six dominant bits, while the passive error flag consists of six recessive bits. An active error flag violates the bit stuffing rules or the fixed-format parts of the frame that is currently being exchanged; hence, it enforces an error condition that is detected by all other stations connected to the network. Each node which detects an error condition transmits an error flag on its own. In this way, as a consequence of the transmission of an error flag, there can be from 6 to 12 dominant bits on the bus.

3.2.2.4 Overload Frame

Overload frame is mentioned for completeness. It is similar to the error frame with regard to the format, and is transmitted by a node that becomes too busy. It is used to create an extra delay between two messages by the slow receivers to slow down operations on the network. Today's CAN controllers are very fast, and so they make the overload frame almost useless.

3.2.3 Medium Access Technique

The medium access control mechanism for CAN network is basically carrier-sense multiple access (CSMA) . When no frame is being exchanged, the network is idle and the level on the bus is recessive. Before transmitting a frame, the nodes have to observe the state of the network. If the network is idle, frame transmission begins immediately; otherwise, node must wait for the current frame transmission to end. Each frame starts with the SOF bit at the dominant level, which informs all the other nodes that the network has switched to the busy state.

Even though very unlikely, it may happen that two or more nodes start sending their frames exactly at the same time. This is actually possible because the propagation delays on the bus, even though very small. Thus, one node might start its transmission

while the SOF bit of another frame is already traveling on the bus. In this case, a collision will occur. In CSMA networks that are based on collision detection, such as, for example, non-switched Ethernet, this unavoidably leads to the corruption of all frames involved, which means that they have to be retransmitted. The consequence is a waste of time and a net decrease of the available bandwidth. In high-load conditions, this may lead to congestion when the number of collisions is so high and then throughput on the Ethernet network falls below the arrival rate, the network becomes stalled.

3.2.3.1 Bus Arbitration

The CAN arbitration scheme allows the collisions to be resolved by stopping the transmissions of all frames involved except the one that is characterized by the highest priority (i.e., the lowest identifier). The arbitration technique exploits the peculiarities of the physical layer of CAN, which conceptually provides a wired-end connection scheme among all the nodes. In particular, the level on the bus is dominant if at least one node is sending a dominant bit; likewise, the level on the bus is recessive if all the nodes are transmitting recessive bits.

When transmitting, each node checks the level observed on the bus against the value of the bit that is being written out. If the node is transmitting a recessive value and the level on the bus is dominant, the node understands it has lost the contention and withdraws immediately. The binary countdown technique ensures that in the case of a collision, all the nodes that are sending lower priority frames will abort their transmissions by the end of the arbitration field, except for the one that is sending the frame characterized by the highest priority (the winning node does not even realize that a collision has occurred). This implies that no two nodes in a CAN network can be transmitting messages related to the same object at the same time. If this is not the case, in fact, unmanageable collisions could take place that, in turn, cause transmission errors. Because of the automatic retransmission feature of the CAN controllers, this will lead almost certainly to a burst of errors on the bus until the stations involved are disconnected by the fault confinement mechanism.

All nodes that lose the contention have to retry the transmission as soon as the exchange of the current (winning) frame ends. They will all try to send their frames again immediately after the intermission is read on the bus. Here, a new collision could take place that also involves the frames sent by the nodes for which a transmission request was issued while the bus was busy. An example that shows the detailed behavior of the arbitration phase in CAN is outlined in Fig. 3.3. Here, three nodes (that have been indicated symbolically as A, B, and C) start transmitting a frame at the same time (maybe at the end of the intermission following the previous frame exchange over the bus). As soon as a node understands it has lost the contention, it switches its output level to the recessive value, so that it no longer interferes with the other transmitting nodes. This event takes place when bit ID 5 is being sent for node A, while for node B this happens at bit ID 2. Node C manages to send the entire identifier field, and then it can keep on transmitting the remaining part of the frame.

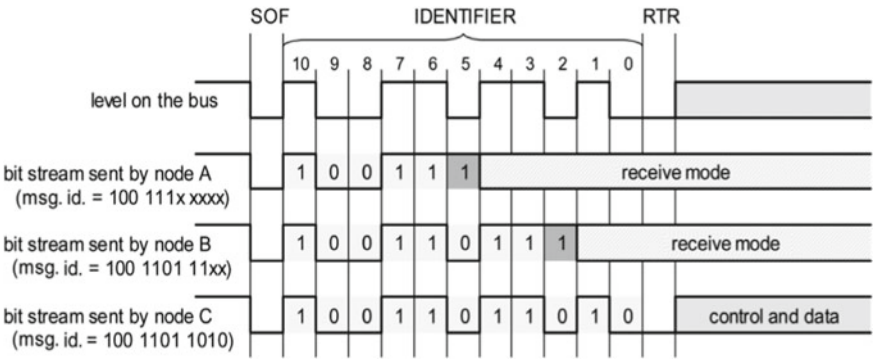


Fig. 3.3 Arbitration phase in CAN

3.2.4 Error Management

One of the main requirements in the definition of the CAN protocol is the need to have a communication system characterized by high robustness, i.e., a system that is able to detect most of the transmission errors. Hence, particular care has been taken in defining error management. The CAN specification foresees five different mechanisms to detect transmission errors:

- *Cyclic redundancy check*: when transmitting a frame, the originating node adds a 15-bit-wide CRC to the end of the frame itself. Receiving nodes reevaluate the CRC to check if it matches the transmitted one. In general, CRC used in CAN is able to discover up to 5 erroneous bits distributed arbitrarily in the frame or errors bursts including up to 15 bits.
- *Frame check*: the fixed-format fields in the received frames can be easily tested against their expected values. For example, the CRC and ACK delimiters as well as the EOF field have to be at the recessive level. If one or more illegal bits are detected, a form error is generated.
- *Acknowledgment check*: the transmitting node checks whether the ACK bit has been set to the dominant value in the received frame. On the contrary, an acknowledgment error is issued.
- *Bit monitoring*: each transmitting node compares the level on the bus against the value of the bit that is being written. When a mismatch occurs, an error is generated. This does not hold for the arbitration field or the acknowledgment slot. Such an error check is very effective to detect local errors that may occur in the transmitting nodes.
- *Bit stuffing*: each node verifies whether the bit stuffing rules have been violated in the portion of the frames from the SOF bit up to the CRC sequence. In the case of when six bits of identical value are read from the bus, an error is generated.

3.2.5 Implementation

According to the internal architecture, CAN controllers can be classified in two different categories: BasicCAN and FullCAN. Conceptually, BasicCAN controllers are provided with one transmit and one receive buffer, as in conventional UARTs. The frame-filtering function, in this case, is generally left to the application programs (i.e., it is under control of the host controller), even though some kind of filtering can be done by the controller. To avoid overrun conditions, a double-buffering scheme based on shadow receive buffers is usually available, which permits a new frame to be received from the bus while the previous one is being read by the host controller. An example of a controller based on the BasicCAN scheme is given by Philips' PCA82C200.

Intel 82526 and 82527 CAN controllers are based on the FullCAN architecture. FullCAN implementations foresee a number of internal buffers that can be configured to either receive or transmit some particular messages. In this case, the filtering function is implemented directly in the CAN controller. When a new frame that is of interest for the node is received from the network, it is stored in the related buffer, where it can then be read by the host controller. In general, new values simply overwrite the previous ones, and this does not lead to an overrun condition.

3.3 Main Features

3.3.1 Advantages

CAN is by far more simple and robust than the token-based access schemes, for example, PROFIBUS when used in multi-master configurations. In fact, there is no need to build or maintain the logical ring, or to manage the circulation of the token around the master stations. In the same way, it is noticeably more flexible than the solutions based on the time-division multiple access (TDMA) or combined-message approaches—two techniques adopted by SERCOS and INTERBUS, respectively. This is because exchanging messages do not have to be known in advance. When compared to schemes based on centralized polling, such as FIP, it is not necessary to have a node in the network that acts as the bus arbiter, which can become a point of failure for the whole system. Since all the nodes are masters in CAN (at least from the point of view of the MAC mechanism), it is very simple for them to notify asynchronous events, such as, for example, alarms or critical error conditions. In all cases where this aspect is important, CAN is clearly better than the above-cited solutions.

With the arbitration scheme, it is certain that no message will be delayed by lower priority exchanges. Since the CAN protocol is not preemptive (as is the case for almost all existing protocols), a message can still be delayed by a lower priority one whose transmission has already started. This is unavoidable in any non-preemptive

system. However, as the frame size in CAN is very small (standard frames are 135 bits long at most, including stuff bits), the blocking time experienced by the very urgent messages is in general quite low. This makes CAN a very responsive network, which explains why it is used in many real-time control applications despite its relatively low bandwidth.

3.3.2 Performances

Since the sampling point is located roughly after the middle of each bit (the exact position can be programmed by means of suitable registers), the end-to-end propagation delay including the hardware delay of transceivers must be shorter than about one quarter of the bit time (the exact value depending on the bit timing configuration in the CAN controller).

As the propagation speed of signals is fixed (about 200 m/μs on copper wires), this implies that the maximum length allowed for the bus is necessarily limited and depends directly on the bit rate chosen for the network. For example, a 250 Kbit/s CAN network can span at most 200 m. Similarly, the maximum bus length allowed when the bit rate is selected as equal to 1 Mbit/s is only 40 m. This, to some degree, explains why the maximum bit rate allowed by CAN specifications ISO1 has been limited to 1 Mbit/s. It is worth noting that this limitation depends on physical factors, and hence it cannot be overcome in any way by advances in the technology of transceivers.

3.3.3 Determinism

CAN is able to resolve in a deterministic way any collision that might occur on the bus because of its nondestructive bitwise arbitration scheme. However, if nodes are allowed to produce asynchronous messages on their own—this is the way event-driven systems usually operate—there is no way to know the exact time of sending a given message. This is because it is not possible to foresee the actual number of collisions a node will experience with higher priority messages. This behavior leads to potentially dangerous jitters, which in several applications, for example those involved in the automotive field, might affect the control algorithms in a negative way and worsen its precision. In particular, it might happen that some messages miss their intended deadlines.

3.3.4 Dependability

Whenever safety-critical applications are considered, where a communication error may lead to damages to the equipment or even injuries to human beings, for example, in automotive x-by-wire systems, a highly dependable network has to be adopted. Reliable error detection should be achieved both in the value and in the time domain. In the former case, conventional techniques such as, for example, the use of a suitable CRC are adequate. In the latter case, a time-triggered approach [10] is certainly more appropriate than the event-driven communication scheme provided by CAN. In time-triggered systems all actions including message exchanges, sampling of sensors, actuation of commanded values, and task activations are known and must take place at precise points in time. In this context, even the presence (or absence) of a message at a given instant could provide significant information (i.e., it enables the discovery of faults).

3.4 Conclusions

CAN is a multi-master serial bus that allows an efficient transmission of data between different nodes. CAN is a flexible, reliable, robust, and standardized protocol with real-time capabilities. Since CAN is message-based and not address-based, it is especially suited when data is needed by more than one location. CAN is ideally suited in applications requiring in a large number of short messages with high reliability in rugged operating environments.

References

1. Johansson KH, Törngren M, Nielsen L (2005) Vehicle applications of controller area network
2. Robert Borsh GmbH (1991) CAN Specification Version 2.0, Part A. Stuttgart, Germany
3. Lawrenz W (1997) CAN system engineering: from theory to practical applications. Springer
4. Farsi M, Barbosa M (2000) CANopen implementation: application to industrial networks. Research Studio Press ltd
5. Robb S (1999) CAN bit timing requirements. Motorola Semiconductor Application Note, AN1798
6. Corrigan S (2002) Introduction to the controller area network (CAN). Texas Instruments Application Report, SLOA101
7. International Organization for Standardization (2003) Road vehicles: controller area network: part 1: data link layer and physical signalling, ISO 11898-1
8. International Organization for Standardization (2003) Road vehicles: controller area network: part 2: high-speed medium access unit, ISO 11898-2
9. International Organization for Standardization (2003) Road vehicles: controller area network: part 3: low-speed, fault-tolerant, medium dependent interface, TC 22/SC 3/WG 1, ISO/PRF 11898-3
10. International Organization for Standardization (2004) Road vehicles: controller area network: part 4: time-triggered communication, ISO 11898-4

Chapter 4
Distributed Control System for Ship Engines Using Dual Fieldbus

4.1 Introduction

Several environmental conditions can cause rapid load changes in a ship engine system, leading to emissions and wasting of resources. In the ship industry, fuel economy and emission reductions are popular issues on account of the current energy shortage and stricter environmental standards. To overcome these problems, the design of a ship engine control system must satisfy hardware and software requirements of the International Association of Classification Societies (IACS).

To design a ship engine control system based on IACS specifications, it is important to understand how communication technology in the industry has been evolving. It has predominantly developed in four stages over the last four decades. The evolution began with direct digital control and transitioned to hierarchical process control. It then evolved to distributed control through near field devices, and finally developed into the distributed control system (DCS) based on the local area network (LAN) and fieldbus. The introduction of DCS with LAN and/or the fieldbus control system has reduced considerable wiring complexity and has made system diagnostics easier and faster [1].

DCS consists of five major components: controllers, input/output (I/O) modules, application software, communication networks, and workstations. These components are combined to deliver an exceptional process automation solution that not only optimizes plant performance and efficiency, but also serves as an extension of management's daily decision-making tools.

Due to the communication network, communication delays or communication losses may occur, which can result in performance degradation or even instability. As a result, researchers have focused on analyzing the network communications systems that are associated with communication delays of DCS [2–4]. An approach proposed in [5] has presented a method to real-time delivery of data in networked control systems to decrease communication delay. In addition to the communication

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_4

delays, in DCS, it is also important that sampled data is transmitted within a sampling period, and that stability of control systems is guaranteed [6]. But, the ship engine distributed control system implies high sampling frequency from sensors, a high probability exists that system entities will fail. The best way to achieve system failure tolerance is by providing redundancy. Redundancy is a common approach to improving the reliability and availability of the system. Adding redundancy increases the cost and complexity of a system design. However, with the high reliability of modern electrical and mechanical components, many applications do not need redundancy to be successful. Nevertheless, if the cost of failure is sufficiently high, redundancy may be an attractive option.

Recent interest in distributed control system redundancy has increased owing to the use of CAN and Modbus for distributed control systems [7]. CAN introduces a message-based protocol designed for automotive applications. Furthermore, it provides the highest speed with reliability and it can be used as a fieldbus on account of the low cost of some controllers and processors [8]. It is also one of the fieldbuses which has better capability of handling electromagnetism disturbances, and it can check errors which are produced in communication bus. Even the distance of the signal communication reached 10 km, CAN still provide digital communication velocity with 50 Kbit/s [9]. CAN is also more excellent than other protocols in many aspects such as capability of real-time delivery, adaption, and security [10].

Modbus is popular industrial protocol being used today for good reasons. It is simple, inexpensive, universal, and easy to use. Even though Modbus has been around since the past century nearly 30 years, almost all major industrial instrumentation and automation equipment vendors continue to support it in new products. Although new analyzers, flow meters and PLCs may have a wireless, Ethernet or fieldbus interface, Modbus is still the protocol that most vendors choose to implement in new and old devices. Another advantage of Modbus is that it can run over virtually all communication media, including twisted pair wires, wireless, fiber optics, Ethernet, telephone modems, cell phones, and microwave. This means that a Modbus connection can be established in a new or existing plant fairly easily. In fact, one growing application for Modbus is providing digital communications in older plants, using existing twisted pair wiring.

Most previous research works have studied control and monitoring only using the CAN [11–14]. A method proposed in [11] improves the entire design level of the ship power system and reduces unnecessary expenses. In [12], an intelligent control system is presented for the ship-hull status. This work promotes ship-hull status monitoring toward the direction of digitization, networking, and intelligence. Moreover, an algorithm for handling real-time message on CAN has been applied to large-scale ship engine network control systems [13].

Furthermore, in [14] a remote control simulation system for a ship engine is presented. It provides a favorable platform for the research of a marine main diesel propulsion control system. In addition, a marine engine remote control system based on distributed processing and dual-redundant CAN network communication technology was proposed for middle-speed and four-stroke marine diesel engine [15]. Its communication system was designed based on the open communication protocol of

CAN, which has the ability to self-test and troubleshoot. The communication system unit has self-checking and troubleshooting functions.

A fault-tolerant CAN controller subsystem was proposed in [16]. It was comprised of three standard CAN controllers and a specifically designed circuit that manages redundancy. The redundancy is provided by a redundancy manager (RM). The RM is the key component of this approach. This circuit performs all the functions relating to redundancy, including typical fault tolerance functions, such as error detection, as well as other functions, such as coordinating the operation of the redundant controllers, even in the absence of faults. However, all these approaches did not consider redundancy of communication link against unexpected failure, which could result damage to the ship system or even a catastrophe.

A network platform for integrated information exchange in the shipboard was proposed in [17]. In that proposal, the standard defined by the International Electrotechnical Commission (IEC61162-4) was adopted as a basic network platform for integrated information exchange. Reference [18] explores the method by which the proportional integrated derivative (PID) controller benefit from smart actuator and fieldbus technologies. A smart actuator scheme suitable for PID controller rearranging was identified and implemented. A CAN bus interface was used for data exchange between the smart actuator and the process controller.

An optimized distributed ship diesel control system based on a master–slave configuration was introduced in [19]. In addition to the many hardware devices attached to the diesel engine, the distributed control system has one master computer and a slave station. The master computer controls the slave station by searching for working information. However, this approach does not consider the communication link redundancy. Reference [20] examines the use of wireless fieldbus in the industrial environment. But, the industrial environment is error prone and the reliability of wireless technology is lower, it is difficult to support a real-time and high data rate transmission.

Existing schemes of DCS techniques did not consider redundancy of communication link against unexpected failure. To overcome the limitations of the existing schemes, a redundant distributed control system (RDCS) for ship engine monitoring is herein proposed. In the proposed approach, the RDCS uses Modbus as a primary communication link and the redundant CAN fieldbus. Redundant CAN provides error detection and retransmission. When an error occurs with Modbus, the recovery process is accomplished by CAN. One of the goals of this recovery is, to restore the coordinated operation that has been lost with the discrepancy. Modbus was used as a primary communication link, because, ship engine interface is developed based on MAN232 which cannot be connected by other fieldbus protocols except Modbus RS232. In ship engine control system, monitoring the whole engine is based on data received from sensors. Due to the importance of precise data about the engine state, sensor data sampling rate is fixed too high which cannot be supported by other fieldbus protocols except Modbus. CAN bus was used as a redundant link, because, CAN bus provides higher speed and faster error recovery than other fieldbus protocols in

case of failure. The main contribution of the proposed RDCS scheme is, it considers communication link redundancy against unexpected failure to provide error detection and retransmission, and to satisfy the recovery timing constraints recommended by IACS. This chapter is an extended version of previous research paper [21].

The remainder of this chapter is organized as follows. Section 4.2 describes the RDCS design scheme. A brief implementation of the proposed RDCS in a real testbed and comparison results are presented in Sect. 4.3. Finally, Sect. 4.4 presents summary of the chapter and the plans for the future work.

4.2 Redundant Distributed Control System

In this section, the proposed RDCS design scheme is presented. Figure 4.1 shows the overall RDCS architecture for a ship engine. The RDCS consists of software (SW) on a personal computer (PC), junction boxes (JB), and sensors that are located on the ship engine. The SW plays the main role in controlling and monitoring. Sensors are placed on the cylinder of the ship engine and are connected to the JBs. Each JB has a micro-control unit (MCU), which controls the functions of data processing, inputting and outputting from/to sensors, and data transmitting using the fieldbus.

As shown in Fig. 4.2, the proposed control system consists of junction boxes connected to each other via a dual fieldbus. Data from sensors are sent to the engine control PC to analyze and monitor the engine parts. This central computer is located in the engine control room. It captures all sensor/signal statuses from the machines and displays the entire condition of the engine room.

The junction boxes convert data from sensors to bit data that is simultaneously readable by the CAN and Modbus. Each junction box MCU is shared by the CAN and Modbus interfaces, as shown in Fig. 4.3.

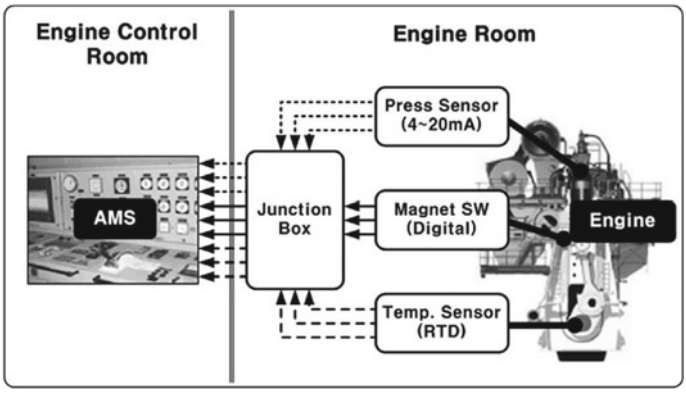


Fig. 4.1 Overall RDCS architecture

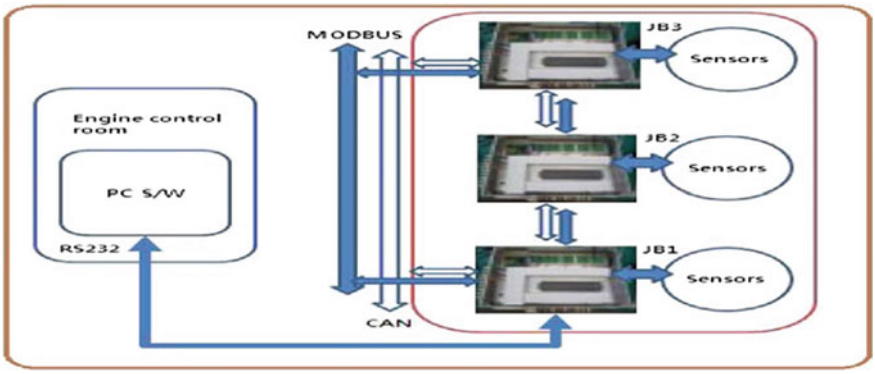


Fig. 4.2 Overview of RDCS with a dual fieldbus

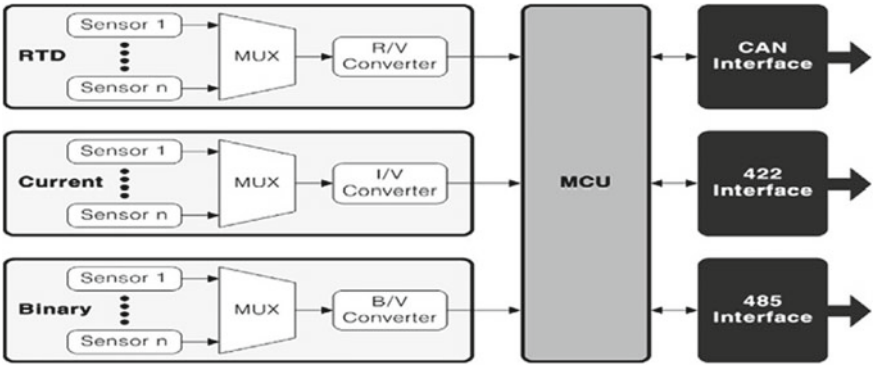


Fig. 4.3 Junction box overview

Figure 4.4 shows the JB initialization. The processes in the junction box are based on interrupt routines from the dual fieldbus and clock (timer). The Modbus interrupt routine is used for sharing real-time IO data. The JB sends a response frame (REP) after receiving a request frame (REQ), as shown in Fig. 4.5. The JB places 0xFF (hexadecimal) into the sensor data to prevent errors, which means that the responses from other JBs do not occur in the delay intervals. The JB sends data when receiving a request from the DCS. Finally, the RDCS recognizes whether the JB is operating or not.

Figure 4.5 shows the procedure of the timer interrupt routine. The time limit used in this study was four in character time. The interrupt routine checks the data update at this time. If no data have been updated, then the earlier data from the JB is used for further error correction.

The timer interrupt routine is reset after receiving a REQ. The REQ is received from a JB via the CAN, which processes the message into the REQ and finally sends a REP to the JB. The CAN interrupt routine function is shown in Fig. 4.5.

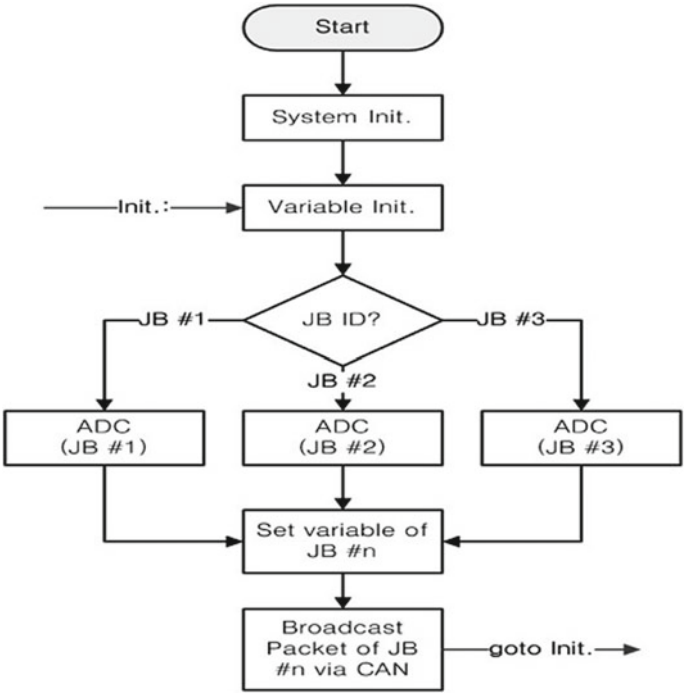


Fig. 4.4 Junction box initialization process

As mentioned above, this approach focuses on redundancy in a ship engine control system for preventing communication link against unexpected failure. The key aspect of the control system is communication between the electronic devices (fieldbus nodes). In this chapter, two fieldbuses: Modbus and the CAN bus are used as communication tools for the ship engine control system. An outstanding feature of the proposed RDCS scheme is the dual fieldbus, which provides redundancy in the case of system failures. To improve the reliability in the alarm monitoring system, two fieldbuses primary Modbus and an alternative CAN bus are used. The main objective of this chapter is to satisfy the recovery time constraint recommended by IACS. When an error occurs during data transmission, the state of the primary Modbus is automatically changed to an inactive state, and that of the CAN bus is changed to an active one.

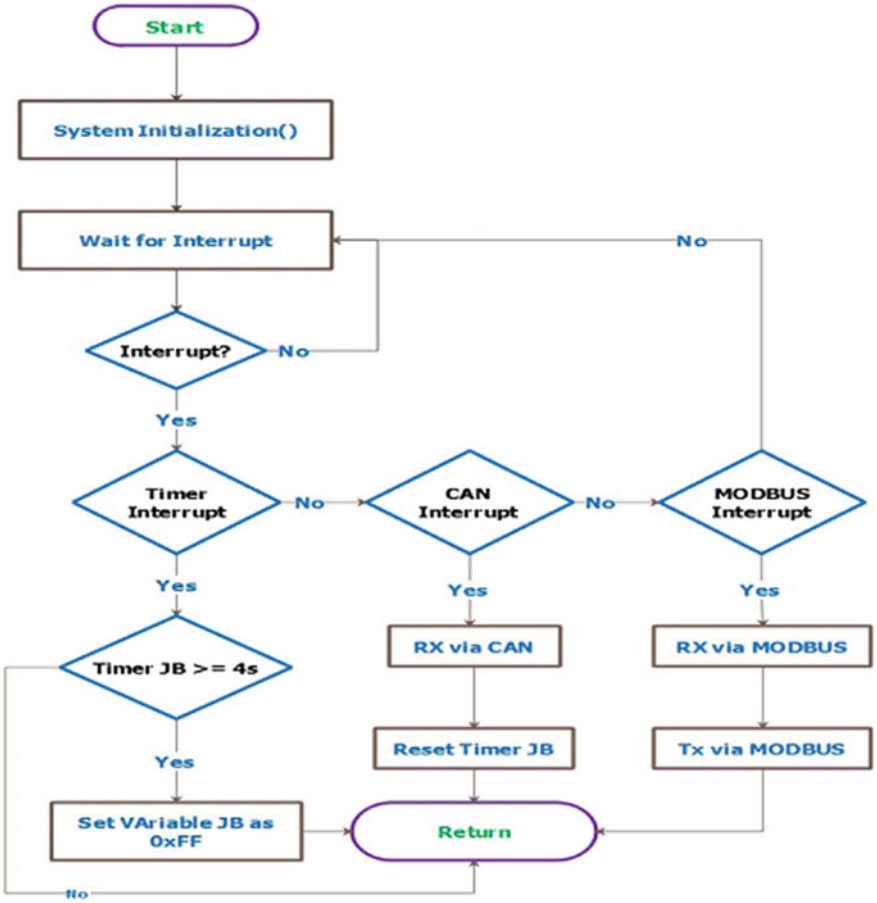


Fig. 4.5 Block diagram for interrupt routines

4.2.1 Modbus Protocol

The Modbus protocol was first published in the 1970s by the American Modicon Company and was used in PLC communication. It is an open standard real-time communication protocol that is widely used in controllers and measuring instruments. More recently, it is becoming an international standard in the industrial automation field. The Modbus protocol supports traditional RS232, RS422, RS485 communication interfaces, as well as the Ethernet interface.

The Modbus protocol has two transmission modes: ASCII and remote terminal unit (RTU). In ASCII mode, the message is expressed by ASCII code and uses a longitudinal redundancy error check. In RTU mode, the message is expressed in binary codec decimal format and uses the cyclic redundant checksum (CRC)

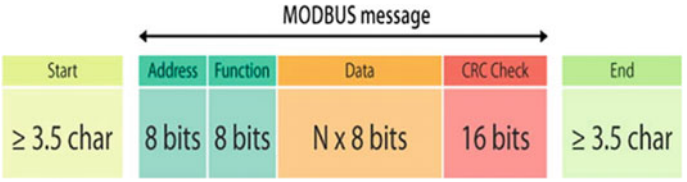


Fig. 4.6 Modbus RTU frame format

Table 4.1 Description of the RTU frame format

Name	Length (Bits)	Description
Start	28	At least 3 1/2 character times of silence (mark condition)
Address	8	Station address
Function	8	The function code
Data	*n* * 8	*n*, number of data bytes
CRC	16	Error checks
End	8	At least 3 1/2 character times of silence between frames

error check. The Modbus RTU is widely used in building management systems and industrial automation systems. This wide acceptance is due in large part to the Modbus RTU ease of use and high communication efficiency [22]. Therefore, considering the above benefits, the RTU transmission mode was selected for the proposed approach.

In the proposed scheme, Modbus is used as the primary fieldbus. It is often used to connect a supervisory computer with an RTU. The Modbus protocol is suitable for systems with high-speed sampling frequencies. The ship engine control system monitors the whole engine based on data received from sensors. Owing to the importance of precise data on the engine state, the sensor data sampling rate is fixed at a level that is too high; thus, it cannot be supported by some fieldbus protocols except Modbus. Modbus RTU is an open serial (RS-232 or RS-485) protocol derived from the master/slave architecture.

Modbus RTU messages are a simple 16-bit CRC. The simplicity of these messages helps to ensure reliability. Because of this simplicity, the basic 16-bit Modbus RTU register structure can be used to pack in floating points, tables, ASCII text, queues, and other unrelated data.

Figure 4.6 shows the Modbus frame format. A Modbus command contains the address of the device for which it is intended. Only the intended device acts on the command, even though other devices might receive it. The basic command can instruct an RTU to change a value in one of its registers. The command is used for control; when the node receives the command, it returns one or more values [23].

Table 4.1 lists the detailed Modbus RTU frame format.

Modbus communication parameters are listed in Table 4.2. Calculation of the total bit time of the Modbus RTU frame depends on the Modbus Function being used. In this case the total bit time is calculated by

Table 4.2 Communication parameters of Modbus

Serial interface	RS485
Protocol	Modbus RTU standard
Baud rates	9600 bps
Data hits	8
Parity	Even
Stop bits	1
Function code	03 (read holding registers)
Master address	01 (AMS)
Slave address	02 (JB1)

$$T_{\mathrm{MD}} = T_{\mathrm{Mbit}} * D_{\mathrm{L}}\ (\mathrm{ms}), \tag{4.1}$$

where T_{Mbit} is the transmission time per each byte calculated by adding all the character bits divided by transmission speed and DL is the frame length of Modbus which is calculated by adding together the request and response bytes.

4.2.2 CAN Protocol

CAN is a serial asynchronous bus used in instrumentation applications for industries such as automobiles. Their digital messages, known as CAN frames, are broadcasted by the nodes on this shared bus through electronic transceivers. The most popular version of the CAN bus has three wires, one ground wire, and two differential CAN signals. The nodes on the bus broadcast the CAN frame when the bus is idle. The receiving nodes acknowledge the receipt of the correct frame by inserting the dominant bit at the acknowledgement bit position. It is possible for two nodes to simultaneously begin transmitting the CAN frame. The node with a lower precedence CAN frame withdraws in the case of a bus contention.

In addition, the CAN specification does not restrict the baud rate to any specific value. There are many popular baud rates. However, all nodes on the bus must operate at the same predetermined fixed baud rate. The maximum allowed baud rate on a CAN bus is 1 Mbps. There is no separate clock signal on the CAN bus to synchronize the node; the CAN frame itself is used for synchronization of the clocks on all the nodes. To effectively achieve this objective, CAN frames have NRZ-5 coding. If there are 5 bits at the same level in the CAN frame, a sixth bit of the opposite level is stuffed by the transmitter. This extra studied bit is removed by the receiver node before processing the CAN frame.

The CAN clock used to sample the bit value of the CAN frame is derived from a clock running at a much higher frequency. The time period of this clock is known as one-time quanta and is denoted by T_{q}. The one bit time of the CAN clock is comprised of many time quanta (T_{q}).

The total CAN clock duration is a sum of the synchronization segment, propagation segment, phase segment 1, and phase segment 2, as given in Fig. 4.7. The synchronization segment of 1 T_q in the clock period is due to the synchronization delay that can occur because the synchronization segment can occur any time within one T_q period.

The propagation segment represents the propagation delay that occurs in the transceiver and cable. Phase segments 1 and 2 are used to handle the phase errors. A synchronization segment is used to synchronize various bus nodes. On transmission, the current bit level is output at the beginning of this segment. If there is a bit state change between the previous bit and current bit, a bus state change is then expected to occur within this segment by the receiving nodes.

A propagation time segment is used to compensate for signal propagation delays on the bus line and through the transceivers of the bus nodes across the network. Phase segment 1 is used to compensate for edge phase errors. This segment may be lengthened during re-synchronization. The sample point is the time at which the bus level is read and interpreted as the value of the respective bit, and its location is at the end of phase segment 1 (between the two phase segments).

Phase segment 2 is also used to compensate for edge phase errors. This segment may be shortened during re-synchronization; however, the length must be at least as long as the information processing time, and it cannot be greater than the length of phase segment 1 [24]. In accordance with the CAN 2.0A specification, the baud rate in this work is set to 250 kbps, and the data length is set to 8 bytes. The total CAN duration T_{Cbit} is made up of the summation of the non-overlapping segments of the nominal bit time in Fig. 4.7 and the synchronization jump width which adjusts the bit clock [25]. Hence T_{Cbit} is given by

$$T_{\text{Cbit}} = T_{\text{syns}} + T_{\text{prs}} + T_{\text{phs1}} + T_{\text{phs2}} + T_{\text{sjw}} * 8, \tag{4.2}$$

where

- T_{Cbit} is the total CAN duration,
- T_{syns} is the first segment in CAN bit timing used to synchronize nodes on the bus,
- T_{prs} is twice the sum of the signal's propagation time on the bus line,

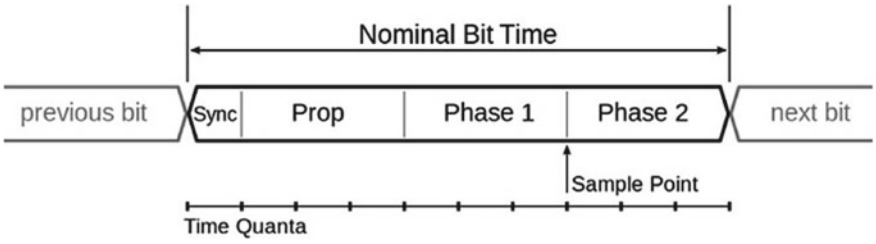


Fig. 4.7 CAN bit timing with 10-time quanta per bit

Table 4.3 Parameters and initial values

T_{syns} synchronization segment	500
T_{prs} propagation time segment	2
T_{ph1} phase segment 1	1
T_{ph2} phase segment 2	2
T_{sjw} synchronization jump width	1

- T_{phs1} and T_{phs2} are the phase segments used to compensate edge phase errors on the bus, and
- T_{sjw} is used to adjust the bit clock as necessary from 1 to 4TQ (time quanta).

The total bit time for CAN specification using Eq. (4.1) can be described as:

$$T_{\text{CD}} = \frac{131}{T_{\text{Cbit}}} \text{ (ms)}, \tag{4.3}$$

where the total data length is 131 bytes in accordance with the CAN 2.0A specifications. Table 4.3 lists the parameters and initial values used in Eq. (4.2).

4.2.3 Redundancy

The ship engine control system is a time-critical one. Therefore, the proposed system design must satisfy requirements predefined by the standards. Requirements are specified by IACS for the DCS design. Table 4.4 lists the requirements for the DCS design in terms of hardware and software.

To meet the IACS design requirements, the proposed control system uses a heterogeneous dual fieldbus. The heterogeneous dual fieldbus consists of the Modbus RTU and CAN. The Modbus RTU is the primary communication media among the electronic devices of the control system. Modbus was chosen as a primary communication link, because, it is compatible with the existing ship engine interface (MAN). When a failure occurs with the Modbus RTU, the recovery process is accomplished

Table 4.4 IACS design requirements

Hardware (HW) requirements	Software (SW) requirements
CPU redundancy	Protocol of International Standard usage
Power redundancy	Isolation of data and power links
HW recovery time	SW response time
Nonvolatile memory	Cyclic redundancy check
Usage	Retransmission assurance

by the redundant CAN. Due to low data rate speed, Modbus cannot be used as redundant communication link. But, CAN has high data rate, and faster error recovery than Modbus. Figures 4.8 and 4.9 illustrate timing in the Modbus RTU channel and the error detection process.

The Modbus RTU message frames are separated by a silent interval of a 3.5 character time (2.96 ms). The inner character silent interval in each frame is set to 1.5 character time (1.25 ms). These features of Modbus are used to detect errors. In this study, it is assumed that, an error occurs when a silent interval is greater than 1.5 character time. When an error occurs, a message frame is declared incomplete and then discarded by the receiver, either the DCS or JB. The total time interval in which an error on each frame is detected is called $t(4.5)$ and is calculated by each electronic device (in this case, the junction boxes), which send data to each other and the monitoring PC

$$t(4.5) = t(1.5) + t(3.5), \tag{4.4}$$

where $t(3.5)$ is the silent interval character time, and $t(1.5)$ is the inner silent interval character time in each frame.

When the junction box cannot receive the Modbus frame after 3.5 character time, then it sends a Modbus REQ message to check its availability and waits for a REP message. If the junction box cannot receive the REP message on time, the CAN bus is switched to maintain the communication process. Figure 4.10 shows the inter-frame delay detection in the CAN bus.

Figure 4.11 describes the redundancy process in the proposed DCS. The time that the redundancy process takes T_{sw} can be defined as follows:

$$T_{sw} = E[T_r + t(3.5) + T_R] + E\left[t(4.5) + I_d + T_q\right], \tag{4.5}$$

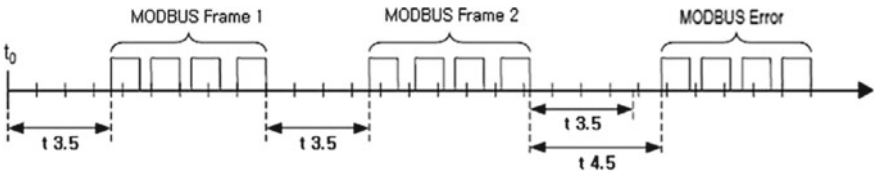


Fig. 4.8 Inter-frame delay detection in Modbus

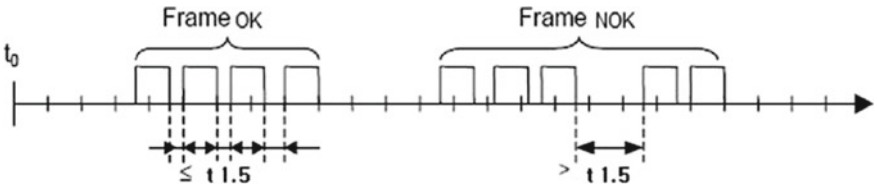


Fig. 4.9 Inter-character delay detection in Modbus

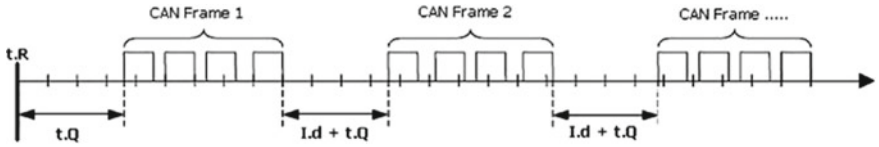


Fig. 4.10 Inter-frame delay detection in CAN bus

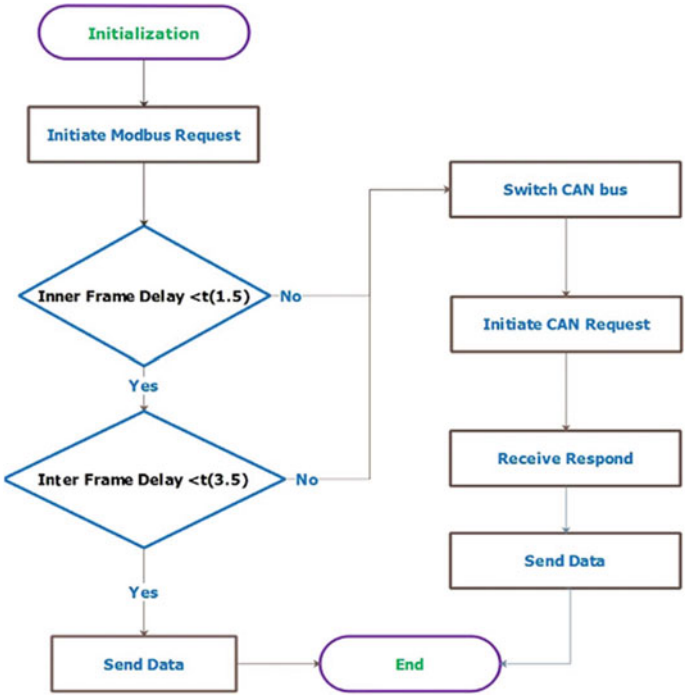


Fig. 4.11 Block diagram of redundancy process

Table 4.5 Parameters and values for computing the switching time

T_r	Send a frame far request to Modbus
T_R	Receive a frame fur response from Modbus
T_q	0.5 ms, at 250 kbps in CAN
I_d	Internal delay for detecting an error

where $E[T_r + t(3.5) + T_R]$ is the expected total time taken to send and receive frame request, and $E\left[t(4.5) + I_d + T_q\right]$ is expected time to identify error and delay. The notations for Eq. (4.5) are given in Table 4.5.

According to IACS, the switching (recovery) time should be less than 2 s. To evaluate the performance of the proposed control system, a testbed was created. It

was determined whether the proposed system satisfies the IACS requirements. In Sect. 4.3, the implementation of the proposed ship engine control system is briefly explained.

4.3 Implementation and Experimental Test

In this section, the implementation and performance evaluation of the proposed DCS is presented. In the implementation, three JBs were created using the CAN controller within AT90CAN128 [26]. A JB has one MCU, ATMEGA AT90CAN128, which is an eight-bit MCU that includes a CAN controller. The CAN controller on the MCU is fully compatible with CAN specifications 2.0A, and 2.0B. It delivers the features required to implement the kernel of the CAN bus protocol according to the OSI reference model. The CAN controller can handle all frame types, including data, remote, error, and overload, while also achieving a bit rate of 1 Mbps [27]. The specifications of the various sensors are listed in Table 4.6.

The analog to digital converter (ADC) converts the analog signal from various resistance temperature detectors (RTDs) into digital values. Figure 4.12 presents a block diagram of the conversion process in JB. The ADC is connected to an eight-channel analog multiplexer that accommodates eight single-ended voltage inputs constructed from the pins of Port F. The single-ended voltage inputs refer to 0 V (GND). The device also supports 16 differential voltage input combinations. The differential input pairs of ADC1 and ADC0, as well as ADC3 and ADC2, are equipped

Table 4.6 Sensor specifications

	Condition	Type	Qty	Range
JB #1	RTD	Temp.	10	0–200 °C
		Temp.	10	−50 to 600 °C
	4–20 mA	Temp.	10	−50 to 600 °C
		Press.	7	0–40 kg/cm^2
	SW	Low	5	Logic
		High	1	Logic
JB #2	RTD	Temp.	15	0–200 °C
	4–20 mA	Press.	1	0–200 kg/cm^2
	S/W	Low	2	Logic
JB #3	RTD	Temp.	8	0–200 °C
		Temp.	1	0–100 °C
		Temp.	18	−50 to 120 °C
	SW	Low	3	Logic
		High	8	Logic

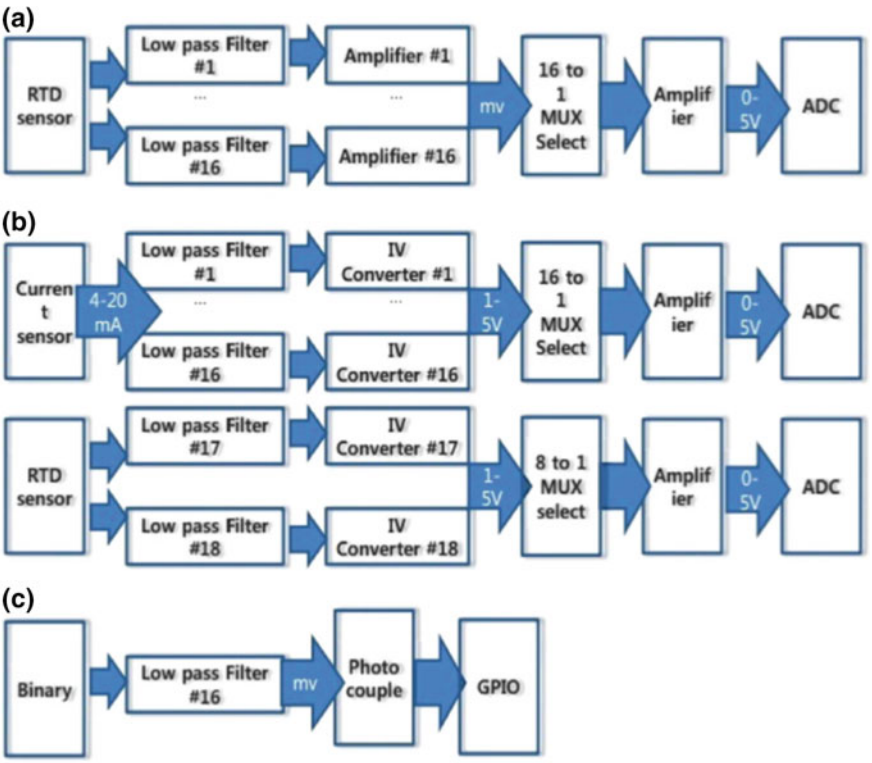


Fig. 4.12 Block diagram of signal conversion system

Fig. 4.13 CAN node overview

with a programmable gain stage, providing amplification of 0 (1×), 20 (10×), or 46 (200×) dB of the differential input voltage before the ADC.

Seven differential analog input channels share a common negative terminal (ADC1), while any other ADC input can be selected as the positive input terminal. If a gain of one or ten is used, 8-bit resolution can be expected. The ADC contains a sample and hold circuit that ensures that the input voltage to the ADC is maintained at a constant level during conversion [28].

An RV converter is used to measure the temperature to convert the register value from the temperature sensor to a voltage from 0 to 5 in amplitude after filtering. The

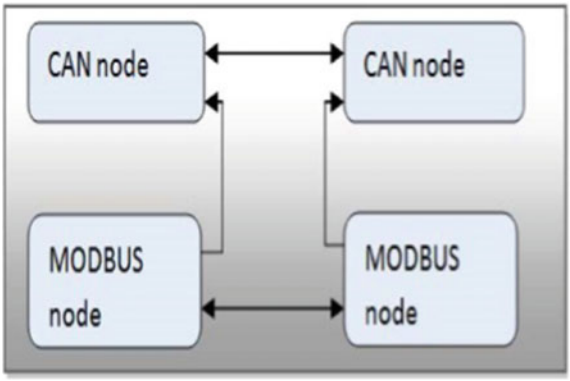


Fig. 4.14 Communication between Modbus and CAN bus

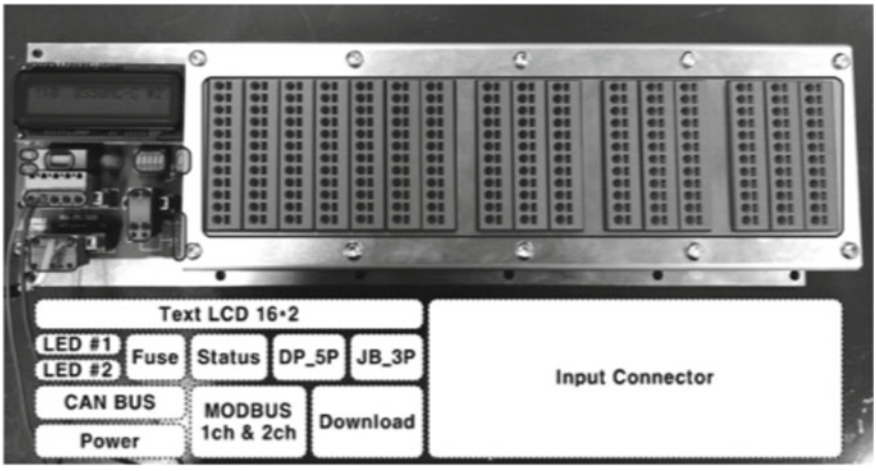


Fig. 4.15 JB overview

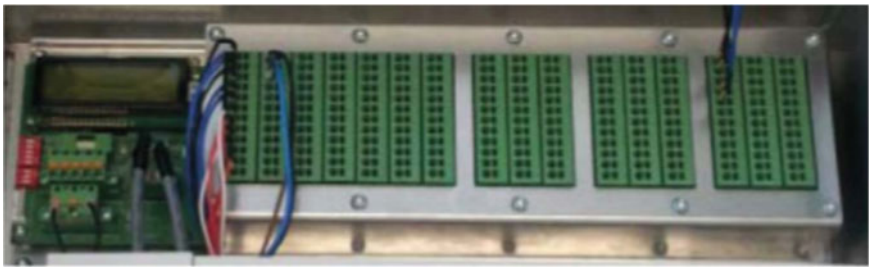


Fig. 4.16 PCB overview

Fig. 4.17 DCS experimental testbed

Fig. 4.18 Experimental results on the screen of the remote monitoring and control device

result from the ADC for the voltage represents a digital value of the sensor data. Each JB consists of two RV converters that support a maximum of 32 inputs. Additionally, RV converters are used to measure the temperature of the thermocouple, the exhaust gas from the cylinders, and the cooling pure water with a pressure sensor. The output values of the sensor are from 2 to 40 mA.

The BV converter indicates the status of the control switch, which can turn on and off, and sensors for detecting cooling oil flow. The sensors for detecting cooling oil flow measure the temperature of the main bearings, the water content of the lubricant in the marine engine system, and the cooling oil level, which causes the signal for the output to turn on and off. The BV converter eliminates the noise across the engine

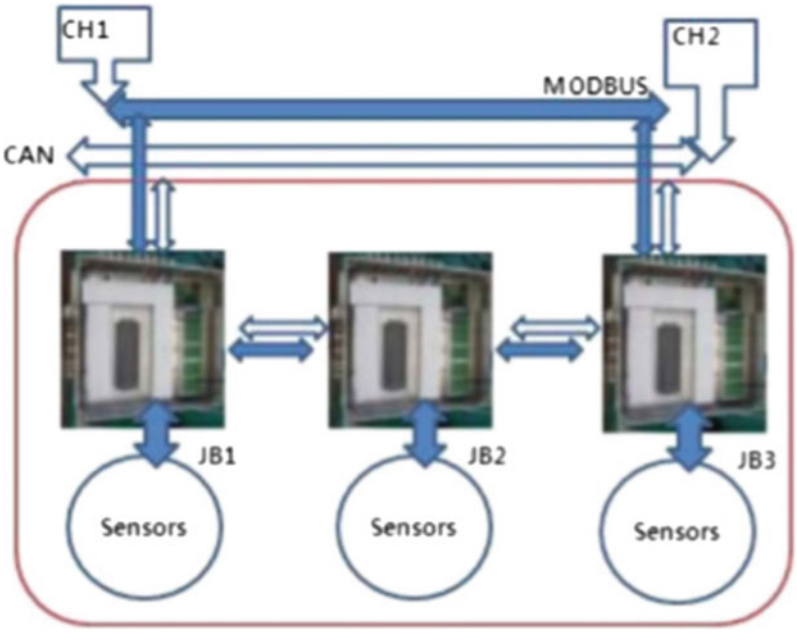


Fig. 4.19 Testbed setup

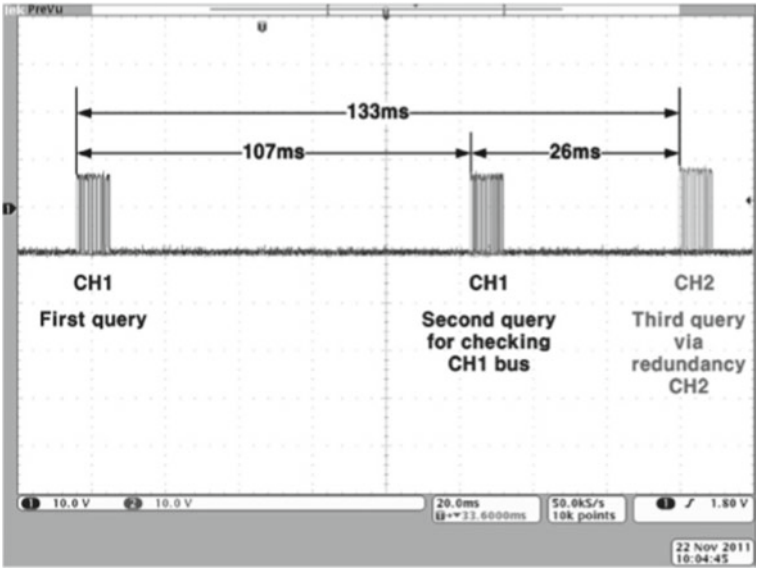


Fig. 4.20 Oscilloscope screenshot during recovery process

Table 4.7 Notations and description

Notations	Description
T_{MD}	Total Modbus bit time
T_{Mbit}	Modbus transmission time
D_L	Modbus frame length
T_{Cbit}	Total CAN duration
T_{syns}	The first segment in CAN bit timing
T_{prs}	CAN propagation time segment
T_{phs1}	CAN phase segment 1
T_{phs2}	CAN phase segment 2
T_{sjw}	Synchronization jump width
T_{CD}	Total bit time for CAN specification
T_r	Frame request to Modbus
T_R	Frame response from Modbus
T_{sw}	Total time of redundancy
I_d	Internal delay for detecting an error
IEC 61162-4	Digital interfaces of a ship with multiple talkers and multiple listeners
RS 232	AN5I/EIA-232 standard for point to point serial communication
RS 422	EIA RS-422-A standard for serial communication
RS 485	EIA-485 standard for serial communication and improvement over RS422
AT90CAN128	Low power CMOS 8-bit microcontroller
ATMEGA	Single-chip microcontroller in megaAVR family
PCA482C250	CAN transceiver chip used in automotive and industrial applications

via a photo coupler after filtering. Arranged data are used for the input of the MCU with general-purpose input output.

The CAN transceiver chip is connected to a JB for I/O data transmission. A Philips PCA82C250 is used as a CAN transceiver chip, as shown in Fig. 4.13. A terminal register 120 Ω is employed to enable the CAN to match the impedance end of the fieldbus. The PCA82C250 is an advanced transceiver product that is used in automotive and general industrial applications with transfer rates up to 1 Mbps. They support the differential bus signal representation described in the international standard for in-vehicle CAN high-speed applications (ISO11898). In this work, the transceiver chip PCA82C250 was used as a communication bridge between the Modbus and CAN, as shown in Fig. 4.14.

The top view of the implemented JB is shown in Fig. 4.15. The processing states are displayed on a liquid crystal display with text. The PCB is completely verified for the product throughout testing, including temperature, humidity, impact, and external disturbance. The overview PCB is shown in Fig. 4.16. The testbed established for the performance evaluation is shown in Fig. 4.17.

The processing results are displayed on the remote PC, as shown in Fig. 4.18. Data exchange is executed from a JB to the PC. In addition, response data arrive within 1 s, even if the input value of the sensor randomly changes. The scenario for the redundancy evaluation is shown in Fig. 4.19. In the scenario, three JBs and three sensors were used. First, the system initialized. Then, input signals in JB 1 and JB 3 were respectively checked with Modbus and CAN. To check system redundancy, the following steps were performed. First, a request frame was sent through Modbus by JB 1, and Modbus was disconnected to cause a system error. A response frame was successfully returned through CAN after redundancy processing.

A request frame was sent through Modbus. Then, an error was created by disconnecting Modbus. A response frame was successfully returned through JB3 after redundancy processing. Figure 4.20 shows the redundancy cycle on an oscilloscope. The fieldbus automatically changes from Modbus to CAN when an error occurs. Redundancy time is required within 2 s according to IACS. The description of the notations used in this chapter is explained in Table 4.7.

4.4 Conclusions

In this chapter, DCS for a ship engine system was proposed. Previous works in this scope considered only a single fieldbus protocol. The existing approaches thus had problems regarding redundancy of communication link against unexpected failures. To overcome the limitations of the existing schemes, the proposed DCS uses Modbus as a primary communication link and redundant CAN bus. When errors occur with the Modbus, redundant CAN bus provides error detection and recovery scheme through retransmission. The proposed DCS is cost effective and flexible for building the control system. Through an experimental evaluation, the proposed design scheme solved the communication problem by switching field buses in reliable way. According to the results, the recovery time is within the time frame required by IACS. Therefore, redundant communication is verified between the DCS and three JBs using the dual fieldbus. In future work, we will verify the reliability of the proposed DCS scheme for more than three JBs.

References

1. Lingqi L, Hanasaki K, Xiangyu W, Yanbin P, Zheng L, Youhua W (1999) Integration of fieldbus into DCS. In: 38th annual conference proceedings of the SICE 1999, pp 1043–1046. http://dx.doi.org/10.1109/SICE.1999.788695
2. Cheng YC, Robertazzi TG (1988) Distributed computation with communication delay (distributed intelligent sensor networks). IEEE Trans Aerosp Electron Syst 24(6):700–712. https://doi.org/10.1109/7.18637

3. Yang F, Wang Z, Hung YS, Gani M (2006) H infinity control for networked systems with random communication delays. IEEE Trans Autom Control 51(3):511–518. https://doi.org/10.1109/TAC.2005.864207
4. Yook JK, Tilbury DM, Soparkar NR (2002) Trading computation for bandwidth: reducing communication in distributed control systems using state estimators. IEEE Trans Control Syst Technol 10(4):503–518. https://doi.org/10.1109/tcst.2002.1014671
5. Kim D-S, Choi D-H, Mohapatra P (2009) Real-time scheduling method for networked discrete control systems. Control Eng Pract 17(5):564–570
6. Kim D-S, Lee YS, Kwon WH, Park HS (2003) Maximum allowable delay bounds of networked control systems. Control Eng Pract 11(11):1301–1313. https://doi.org/10.1016/s0967-0661(02)00238-1 (URL http://www.sciencedirect.com/science/article/pii/S0967066102002381)
7. CAN Protocol Specification (2013). http://www.can-cia.org. Accessed 13 Feb 2015
8. Andrn F, Strasser T, Zoitl A, Hegny I (2012) A reconfigurable communication gateway for distributed embedded control systems, pp 3720–3726. https://doi.org/10.1109/iecon.2012.6389299
9. Lin Qing YX et al (1999) Field bus and network integration, test control technique
10. Ran P, Wang B, Wang W (2008) The design of communication convertor based on CAN bus, in: IEEE International Conference on Industrial Technology, ICIT 2008, pp 1–5. http://dx.doi.org/10.1109/ICIT.2008.4608607
11. Chen D, Xia L, Wang H (2008) Modeling and Simulation of Monitor—Control Network in Ship Power Station, pp 384–388. http://dx.doi.org/10.1109/PEITS.2008.97
12. Shenhua Y, Minjie Z, Xinghua W, Chunsen C (2010) Design and implementation of intelligent monitoring system for ship-hull status based on CAN bus 1 384–388. http://dx.doi.org/10.1109/ICOIP.2010.217
13. Noh DH, Kim DS (2014) Message scheduling on can bus for large-scaled ship engine systems, IFAC Proc. 47(3):7911–7916
14. Kay S, Michels J, Chen H, Varshney P (2006) Reducing probability of decision error using stochastic resonance, IEEE Signal Process. Lett. 13(11):695–698. http://dx.doi.org/10.1109/LSP.2006.879455
15. Cao H, Ma J, Zhang G, Zhang J, Ren G (2010) Marine main engine remote control system with redundancy can bus based on distributed processing technology, in: 2010 International Conference on Intelligent Control and Information Processing, ICICIP, pp 638–640. http://dx.doi.org/10.1109/ICICIP.2010.5564343
16. Guerrero C, Rodriguez-Navas G, Proenza J (2002) Hardware support for fault tolerance in triple redundant can controllers, in: 9th International Conference on Electronics, Circuits and Systems, vol 2, pp 457–460. http://dx.doi.org/10.1109/ICECS.2002.1046195
17. J K-W, L J-W, P J-H, K SY, H-C Park, J-S Lee (2010) Development of network platform for integrated information exchange on shipboard, in: IJCSNS Int J Comput Sci Netw Secur 10 Jan 2010
18. Lee D, Allan J, Thompson HA, Bennett S (2001) PID control for a distributed system with a smart actuator, Control Eng Pract 9(11):1235–1244. http://dx.doi.org/10.1016/S0967-0661(01)00069-7 (pID control. URL http://www.sciencedirect.com/science/article/pii/S0967066101000697)
19. Chang-kun H (1996) A distributed control system of ship diesels, in: Proceedings of the IEEE International Conference on Industrial Technology, ICIT'96, pp 1–5. http://dx.doi.org/10.1109/ICIT.1996.601528
20. Choi D, Kim DS (2008) Wireless fieldbus for networked control systems using lr-wpan, Int J Control Autom Syst 6(1):119
21. Gereziher WA, Dong-Seong K (2017) Distributed control system for ship engines using dual fieldbus, Computer Standards & Interfaces, vol 50, pp 83–91
22. Wang K, Peng D, Song L, Zhang H (2014) Implementation of modbus communication protocol based on arm coretx-m0, in: 2014 IEEE International Conference on System Science and Engineering, ICSSE, pp 69–73. http://dx.doi.org/10.1109/ICSSE.2014.6887907

23. Modbus Protocol Specification (2013) http://www.can-cia.org⟩ (accessed 2/3/2015)
24. Lee H, Yi DK, Lee JS, Park G-D, Lee JM (2010) Marine engine state monitoring system using distributed precedence queue mechanism in CAN networks, in: Proceedings of the Third International Conference on Intelligent Robotics and Applications, ICIRA 2010, Shanghai, China, November 10–12, 2010, Part I, Springer, Berlin, Heidelberg, pp 237–245. http://dx.doi.org/10.1007/978-3-642-16584-9_22
25. D.C.W.A. Engineer for Transceivers (2014) Can Bit Timing to Optimize Performance, vol 48, September 2014 (URL http://www.analog.com/library/analogdialogue/archives/48-09/CAN_bit_timing.html)
26. At90can128 Data Sheet (2013) http://www.atmel.com⟩ (accessed 7/3/2015)
27. Lee J, Lee J (2010) Scheduling of can-based network for marine engine state monitoring, in: The 7th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI 2010
28. Modbus Over Serial Line (2013) http://www.modbus.org. (accessed 20/5/2015)

Chapter 5
Implementing Modbus and CAN Bus Protocol Conversion Interface

5.1 Introduction

Any embedded system generally consists of one or more micro-processors or micro-controller and a number of peripherals IC's like EEPROM, real-time clock (RTC), watchdog timer and sensors, etc. In communication system design, a key challenge is the ability to make different components from different manufactures communicates with each other. A number of field buses are available to exchange the serial data among one or more controllers and a number of field devices that are communicating with each other. However, fieldbus standards are currently not uniform, which brings many difficulties in system design, as different equipment from different manufacturers follow different standards. For a reliable system design, there is a need of efficient communication interface to make the communication possible. Many serial communication protocols like RS-232/RS-485, I2C, SPI, Modbus and CAN bus, etc., were used in embedded systems. All these protocols have their own advantages and limitations. Generally, different manufactures follow different protocols and standards. This makes the system integration task very difficult. Hence, there must be several means to make this task easier. Protocol conversion interface is one of the possible solutions for this problem. CAN bus and Modbus are two most common fieldbus protocol used in industrial control systems. This chapter implements a CAN bus to Modbus protocol conversion interface. Both sides of serial connections are isolated to provide perfect protection against lightning, surges, high-voltage transients. After the brief introduction of each protocol, this chapter briefly explain the hardware and software design of CAN bus to Modbus protocol conversion interface. Figure 5.1 describes the basic overview of the system and how different devices are connected in the system [1].

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_5

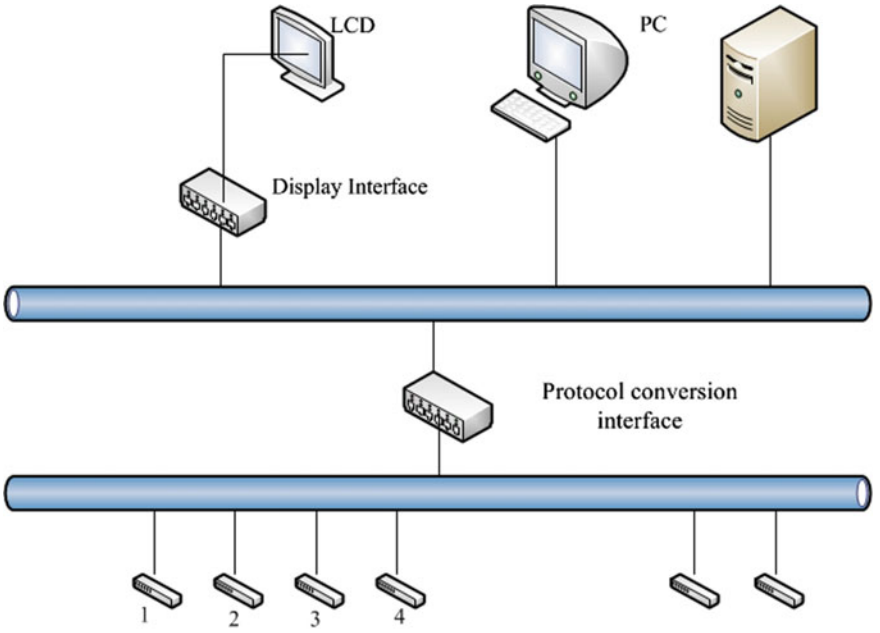


Fig. 5.1 System structure

5.2 Modbus and CAN Bus

5.2.1 Modbus

5.2.1.1 Overview

Modbus is a communication protocol widely used in distributed control applications. Modbus was initially introduced in 1979 by Modicon (a company now owned by Schneider Electric) as a serial-line protocol for communication between "intelligent" control devices. It has become a de facto standard implemented by many manufacturers and used in a variety of industries. Modbus is a master–slave communication protocol which describes the process a master uses to request an access to slave, and how the slave will respond to these requests, and how errors will be detected and reported. Master can initiate transactions (called "queries") and slave respond by supplying the requested data to the master, or by taking the action requested in the query. The master can address individual slaves, or can initiate a broadcast message to all slaves. Slaves return a message (called a "response") to queries that are addressed to them individually. Responses are not returned to broadcast queries from the master. Figure 5.2 describes the master–slave query-response cycle.

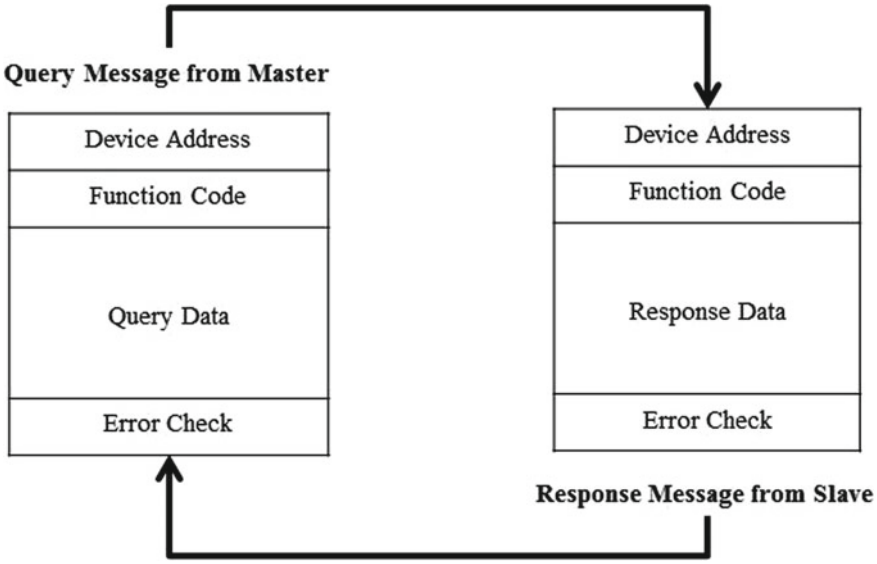


Fig. 5.2 Modbus master–slave query-response cycle

5.2.1.2 Message Frame Format

The Modbus serial-line specifications describe physical and link-layer protocols for exchanging data [2]. Two main variants of the link-layer protocol are defined and two different types of serial lines are supported. In addition, the specifications define an application layer protocol, known as the Modbus Application Protocol, for controlling and querying devices [3]. The application protocol was originally intended for devices connected via the Modbus serial-line protocol and was summarized as following:

- The application protocol follows the same master–slave design as the serial-line protocol in Fig. 5.3. Each transaction at the application layer is a simple query-response exchange initiated by the master node and addressed to a single device. Both requests and responses fit in a single serial-line frame.
- Each device on the same serial line has an 8 bit address. Addresses 0 and 248 to 255 are reserved. There can be at most 247 devices on a single line.
- The maximal length of a Modbus frame is 256 bytes. One byte is the device address and two bytes are used for CRC. The maximal length of a query or response is 253 bytes.

Subsequently, Modbus specifications were extended to support other types of buses or networks. Modbus Application Protocol assumes an abstract communication layer that allows devices to exchange small packets. Serial-line Modbus remains an option for implementing this communication layer, but other networks and protocols may be used. Increasingly, TCP/IP is being used as the communication

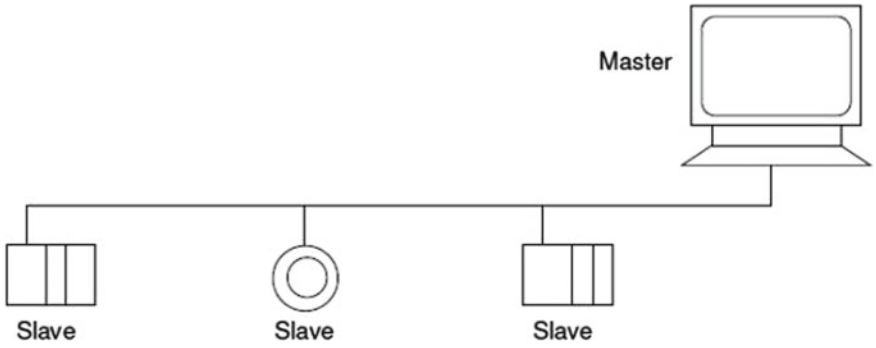


Fig. 5.3 Modbus over serial line: master/slave architecture

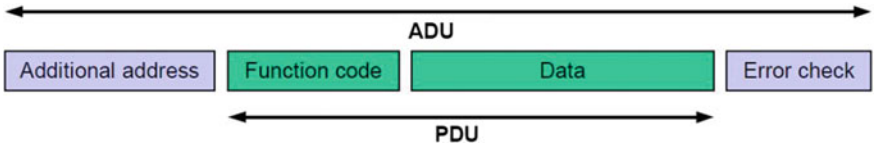


Fig. 5.4 Modbus date format

layer for Modbus. Modbus over TCP/IP specification [4] describes how to implement the Modbus communication layer using TCP.

Master's query consists of slave device (or broadcast) address, a function code defining the requested action, any data to be sent, and an error-checking field. The slave's response contains fields confirming the action taken, any data to be returned, and an error-checking field. Figure 5.4 shows the PDU and ADU of Modbus protocol.

Two different serial transmission modes are defined: The RTU mode and the ASCII mode. It defines the bit contents of message fields transmitted serially on the line. It determines how information is packed into the message fields and decoded. The data transmission rate of ASCII mode is a little lower than RTU mode. So, when need to send large data, user always uses RTU mode. The standard Modbus protocol is to use a RS-232C compatible serial interface, which defines the port pin, cable, digital signal transmission baud rate, parity.

5.2.2 CAN Bus

CAN is a serial communication protocol which was originally developed in February 1986, by Robert Bosch GmbH mainly for applications in the automotive industry but also capable of offering good performance in other time-critical industrial applications. The CAN protocol is optimized for short messages and uses a CSMA/Arbitration on Message Priority (CSMA/AMP) medium access method.

Thus the protocol is message-oriented, and each message has a specific priority that is used to arbitrate access to the bus in case of simultaneous transmission. The bit-stream of a transmission is synchronized on the start bit, and the arbitration is performed on the following message identifier, in which a logic zero is dominant over a logic one.

A node that wants to transmit a message waits until the bus is free and then starts to send the identifier of its message bit by bit. Conflicts for access to the bus are solved during transmission by an arbitration process at the bit level of the arbitration field, which is the initial part of each frame. Hence, if two devices want to send messages at the same time, they first continue to send the message frames and then listen to the network. If one of them receives a bit different from the one it sends out, it loses the right to continue to send its message, and the other wins the arbitration. With this method, an ongoing transmission is never corrupted.

In a CAN-based network, data are transmitted and received using Message Frames that carry data from a transmitting node to one or more receiving nodes. Transmitted data do not necessarily contain addresses of either the source or the destination of the message. Instead, each message is labeled by an identifier that is unique throughout the network. All other nodes on the network receive the message and accept or reject it, depending on the configuration of mask filters for the identifier. This mode of operation is known as multicast.

The CAN communications protocol, ISO-11898, describes how information is passed between devices on a network and conforms to the open systems interconnection (OSI) model that is defined in terms of layers. Actual communication between devices connected by the physical medium is defined by the physical layer of the model [5]. The International Standards Organization (ISO) defined a standard ISO 11898 which incorporates the CAN specifications to meet some of the requirements in the physical signaling, which includes bit encoding and decoding (Non-Return-to-Zero, NRZ) as well as bit timing and synchronization. Using serial bus network mechanisms, the existing CAN applications send messages over the network. In the CAN systems, there is no need for central controller as every node is connected to the every other node in the network. CAN communications protocol, ISO 11898:2003, describes how information is passed between devices and conforms to OSI model that is defined in terms of layers. Actual communication between devices by the physical medium is defined by the physical layer of the model. The ISO 11898 architecture defines the lowest two layers of the seven layer OSI/ISO model as the data link layer and physical layer as shown in Fig. 5.5.

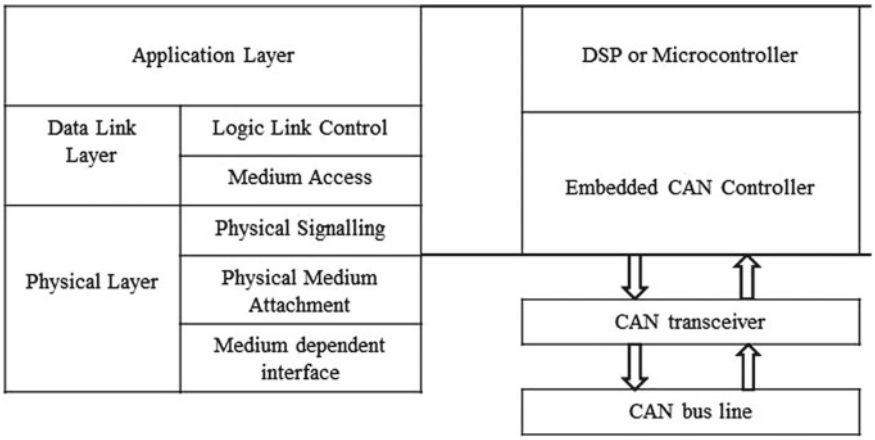


Fig. 5.5 The layered ISO standard 11898 architecture

5.3 Conversion Interface Design

5.3.1 Hardware Design

Design of this CAN bus to Modbus protocol conversion interface is done by using PIC32MX-XXX series microcontroller. This series of microcontrollers has 6 UART and 2 CAN modules. These on chip modules are appropriate of the design of protocol conversion interface. This design also included a CAN transceiver (ISO-1050) and As the Modbus using the RS-485 serial interface, this design uses RS485 transceiver

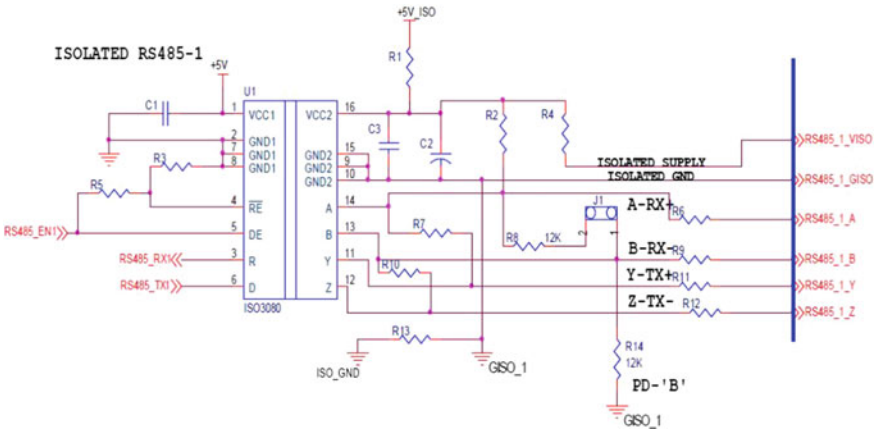


Fig. 5.6 Schematic diagram for Modbus communication

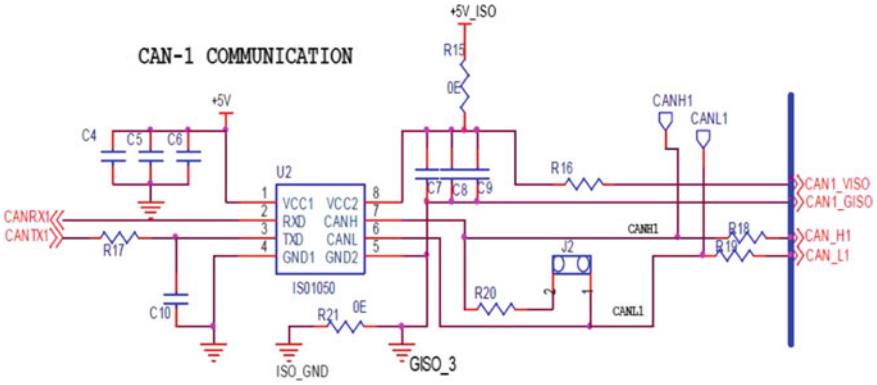


Fig. 5.7 Schematic diagram for CAN bus communication

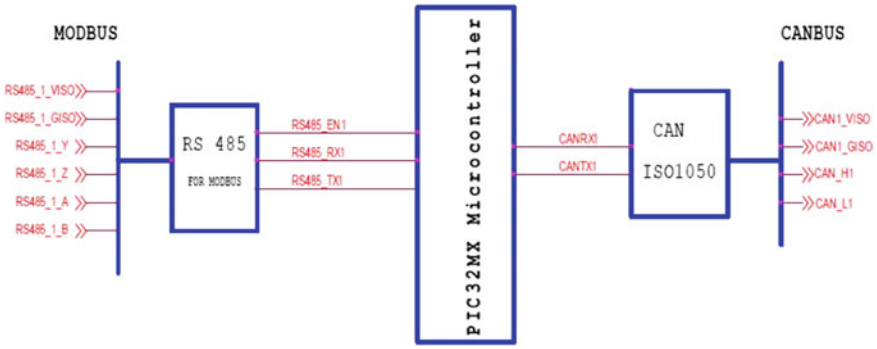


Fig. 5.8 Conversion interface

(ISO-3050). The ISO-1050 is an isolated CAN transceiver that meets or exceeds the specifications of the ISO-11898 standard [6] and ISO-3080 is an isolated full-duplex differential line drivers and receivers.

This CAN transceiver provides differential transmit capability to the bus and differential receive capability to a CAN controller at signaling rates up to 1 megabit per second (Mbps). The internal oscillator circuit of PIC32 microcontroller is used to generate the system clock. This system is working on a high frequency of 32 MHz. Block diagram of protocol conversion interface is shown in Fig. 5.8. Figure 5.6 shows the schematic diagram for Modbus communication using RS-485 and Fig. 5.7 shows the schematic diagram for CAN bus communication.

5.3.2 Software Design

This protocol conversion interface is working on master–slave technique and only master can initiate the communication. Here, CAN bus is chosen as the master and

Modbus as slave. Only CAN bus can start the communication and Modbus devices can respond according to request made by the master. The entire communication is controlled by an event-driven interrupt. Whenever CAN master wants to communicate with a device connected over the Modbus, it will then interrupt the communication. On receiving this interrupt, CPU enters into an interrupt service routine (ISR). In this routine, CPU receive the data from CAN master and checks the integrity of the data using CRC check. If data is found to be valid then program enters into a routine which convert this data into Modbus format. This format contain slave address, function code, data field, and CRC field. Thus, this conversion interface encapsulates the data in Modbus protocol format to send to the Modbus site. Now, slave device receives the data and responds back with the response data. Conversion interface receives this response data. After receiving message conversion interface analysis data, and then convert to CAN protocol format, send to master. Note that, due to the length of CAN bus data transmission up to 8 bytes, if Modbus protocol transmits data is longer than 8 bytes it will send data many times.

5.4 Conclusions

This chapter presented a survey on Modbus and CAN bus and the designing of CAN bus and Modbus protocol conversion interface. By using the hardware and software logic as explained above CAN bus to Modbus protocol conversion interface is implemented. There are a number of advantages of using fieldbus communications when adding or replacing a remote control for your rectifier. It will save installation costs, is easier to troubleshoot, and allows you to remotely program your process. It gives the ability to configure, monitor, and troubleshoot a rectifier over a great distance. This new generation of advanced remote controls gives a user the ability to select a specific controller for a specific process, with many standard options already integrated.

References

1. Guohuan L, Hao Z, Wei Z (2009) Research on designing method of can bus and modbus protocol conversion interface. In: International conference on future biomedical information engineering, pp 180–182, Dec 2009
2. Modbus over serial line: specification and implementation guide V1.0. Available at http://www.modbus.org, Dec 2002
3. Modbus application protocol specification V1.1a. Available at http://www.modbus.org, June 2004
4. Modbus messaging on TCP/IP: implementation guide V1.0a. Available at http://www.modbus.org, June 2004
5. Boterenbrood H (2000) Canopen high-level protocol for CAN-bus. NIKHEF, Amsterdam, 20 Mar 2000
6. Tindell K, Burns A, Wellings A (1995) Calculating controller area network (can) message response times. Control Eng Pract 3(8):1163–1169

Chapter 6
MIL-STD-1553 Protocol in High Data Rate Applications

6.1 Introduction

MIL-STD-1553 is a classical digital bus interface originally developed by the U.S. air force specifically for military and avionics applications [1–3], but later on, it has been opened for commercial applications. Currently, the standard has been widely used by automation and avionics industries, due to its robustness, highly reliable data transfer, and a strong market leader in command and control applications for the last 30+ years [3–5]. Furthermore, the pervasiveness of MIL-STD-1553 is making its way to adopt new applications in the future ranging all the way from military to space, avionics, command, and control systems, providing high degree of interoperability [4], and consistent protocol reliability. Airbus is the world's largest aircraft manufacturer that has adopted the same standard for its modern flight control system in A350 XWB aircraft [4], due to its reliability, robustness, and having 30+ years of proven flight control experience. The objective of fieldbus in automation industry is to reduce the number of wiring and reconfiguration at low cost [6]. Bringing 802.11 wireless LAN into automation industry leads to the concept of "wireless fieldbus", but the legacy standard did not guarantee the reliability, timeliness, and flexibility [6, 7]. The TCP/IP is an open communication standards, often preferable in flexible data routing and reliable data transmission [8, 9] which do not guarantee the communication routing reliability in MIL-STD-1553. The main reason is that MIL-STD-1553 network architecture uses bus controller for routing and data control.

MIL-STD-1553 characteristics like high reliability, availability, fault tolerance, and interoperability have made it a superior choice and well-suited solution in avionics industry [3, 7, 9–11]. The Integrated Avionics Systems (IAS) is a complex flight control system in modern fighter being able to collect, code, distribute, and store air duty information during critical mission. MIL-STD-1553 provides efficient, low cost, and lightweight means of multiplexing computers in modern military avionics systems. These devices will communicate with aircraft internal subsystem efficiently and reliably. According to [12], the wide acceptance of 1553 has become recognized

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_6

as a general purpose bus adopted by various automation industries, besides avionics for multiplexing computers. This standard is controlled by Automotive Society of Engineer (SAE), continuously enhancing the standard performance for future military avionics systems.

6.2 Related Works

MIL-STD-1553 is a digital time division multiple data bus primarily developed by the U.S. air force [1, 5, 7, 9, 13] since its inception in 1973, and consciously under revision for next-generation avionics systems as well as commercially available for applications like international space stations, missiles, tanks, ships, and satellites [1, 3, 5, 7–9, 13]. Size and weight are the most critical factors in avionics industry [1], and MIL-STD-1553 has been solving this challenging issue by minimizing the point-to-point aircraft internal interconnection for the last 45 years, introducing the real time flexibility. The success of digital fieldbus in automation and avionics industry is mainly due to reduction of wiring (size, weight, power, and cost) and the standard is continuously upgrading to completely eliminate the internal wiring, by introducing wireless connectivity among components. Wireless medium (WLAN+ others wireless technologies) are strictly limiting the system reliability, efficiency, and sensitive high-speed data communication in critical military, space, and avionics applications [6]. IP core is a single intelligent, versatile chip capable of to implement and control all of the three bus functionalities like initiating commands/response by Bus Controller (BC), acknowledgement made by Remote Terminal (RT), and status reporting by Bus Monitor (BM) [7, 8, 12, 14]. The data traffic of IAS [14] depends on MIL-STD-1553B bus. The simulation result conducted using HDL (ALDEC) shows that IP core has many advantages including error detection capability making it highly reliable along with MIL-STD-1553.

In [8, 9], the IP datagram packets are encapsulated in MIL-STD-1553B data message using IP over MIL-STD-1553 System (IPo1553 System), having backward compatibility to legacy 1553 network. According to IPo1553, some of the devices enabled by IPo1553 send and receive IP datagrams while others continue to transmit the data using legacy 1 Mbps standard. The IPo1553 provides an additional communication option to continuously deliver data using IP protocol along with legacy 1553 standard. Paper [7] introduces hardware and software interface codesign using BU-61580 and 32-bit SPARC V8 processor to achieve high reliability data communication between 1553 Remote Terminal and Bus controller. Mass, volume, and power consumption are three most important design factors for MIL-STD-1553 in aerospace applications [10]. The author introduces the design of 1553B bus low power technologies, aiming to minimize the power consumption to improve data communication efficiency and reliable data transfer. Performance analysis for three low power technologies namely, 1553B protocol based on SOC chip, RS485 transceiver, and infrared are compared and analyzed for mass, volume, and power consumption so that each low power technology is well suited for specific

applications. According to [10], the low power 1553B technology will be used in one of the Chinese space mission and future satellites.

The degree of testing and simulation requirements of 1553 serial bus for efficiency and data communication reliability is analyzed in [12]. The proper approach with optimal trade-offs between total flexibility and cost must be evaluated to select proper bus testing technique. Paper [15] proposes performance evaluation of high-speed data rate over MIL-STD-1553 bus using Discrete Multitone Technology (DMT) on existing 1553 standard. The system architecture uses FPGA transmitter and receiver for verification and experimental assessment. The transmitted and received data files of 250 MB (3500 bursts) and 1.4 GB (20,000 bursts) using 100 Mbps data rate over MIL-STD-1553 bus are analyzed. Performance evaluation of designed system improves the transmission of large volume of data continuously using high-speed data rate. The improved high-speed transmission technique with bit error rate of 10^{-9} signifies the reliability of large volume of data transfer using 100 Mbps compared to traditional 1 Mbps system.

MIL-STD-1553 feasibility study on network capacity analysis has been conducted in [4] by Data Device Corporation (DDC), implementing the legacy system for high-speed data communication resulting lower Bit Error Rate (BER). According to [4], the actual high-speed data communication performance and bus reliability depends on many factor such as bus length, stub length, number of devices connected to bus, waveform signaling, and data encoding scheme.

6.3 MIL-STD-1553 Network Protocol Infrastructure

MIL-STD-1553 network architecture is the aircraft internal time division command/response multiplex data bus, widely accepted standard by military and avionics currently running version B [8, 12, 13].

6.3.1 MIL-STD-1553 Hardware Elements

MIL-STD-1553 network architecture can be divided into four main components as described in [2, 3, 13, 16].

- Bus Controller (BC)
- Remote Terminal (RT)
- Bus Monitor (BM)
- Transmission media/1553 bus cable.

The MIL-STD-1553 network architecture is depicted in Fig. 6.1 including BC, RT, BM, primary bus, and optional secondary bus acting as a backup bus.

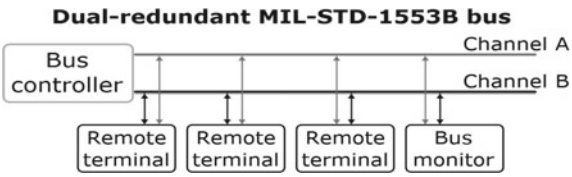


Fig. 6.1 MIL-STD-1553 network architecture

6.3.1.1 Bus Controller

Bus Controller (BC) is responsible to initiate the command/response communication [7, 8, 12, 17] over digital data bus. BC happened to be more than 1 in number, but only one active BC is operational at any given time [8]. The remote terminals and other system components connected to data bus are directed to receive commands initiated by BC and response back as specified in message. The commands are either related to data management/traffic control on data bus or directed to subsystems (terminals) to response back as specified in the message. BC uses command/response methodology to communicate over the bus acting as a bus regulator in MIL-STD-1553 communication interface [5, 12].

6.3.1.2 Remote Terminal

The primary function of Remote Terminal (RT) is to send/receive communication messages from BC, among RTs and subsystem components over the bus and up to 31 RTs can communicate over MIL-STD-1553 digital bus as in Fig. 6.1. RT receives and decomposes messages directed from BC and RTs to respond accordingly. RT is capable to detect transmission errors, performs the data validation tests once the data is received and report data transmission status (failed/acknowledged/delayed). RT is capable to initiate a response strictly in 12 μs, once it is directed from BC although it is not defined in the standard [8]. According to MIL-STD-1553, RT can be any computer/component/control system connected to bus like aircraft navigational unit, air traffic control computer, weapon control unit, or even simple PC connected to bus can be considered as RT [12, 17].

6.3.1.3 Bus Monitor

Bus Monitor (BM) is one the most important elements of MIL-STD-1553 network architecture with primary function are to collect and monitor the data transmission over data bus. BM can be used for "off-line application" like flight test recording, maintenance recording, and mission analysis [8, 17]. BM is also considered as bus traffic recorder and error detector.

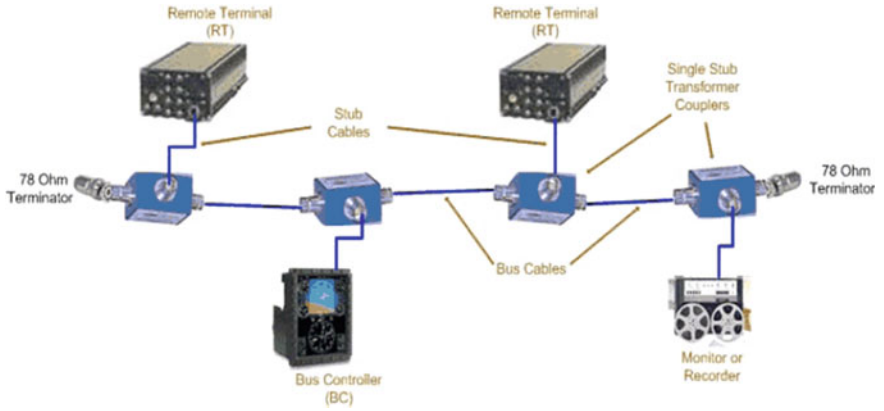


Fig. 6.2 MIL-STD-1553 network infrastructure real time

Table 6.1 Summary of data bus requirements

Application	DOD avionics
Data rate	1 MHz
Word length	20 bits
Number of data bits/word	16
Transmission technique	Half duplex
Operation	Asynchronous
Encoding	Manchester II bi-phase
Bus coupling	Transformer
Bus control	Single or multiple
Transmission media	Twisted pair shielded

6.3.1.4 Transmission Media/1553 Bus Cable

Twisted shielded pair cable is acting as digital transmission bus connecting BC, RTs, and BM through stub cables along with signal terminator of 78 Ω at both ends. Transformer coupling and isolation resistors are introduced at terminal and on bus to reduce noise and shorts [17]. The twisted pair cable acting as a digital data bus preventing noise cancelation and shielding provides protection from external electromagnetic interference, thus ensuring the data integrity [1]. Maximum number of RTs connected to a single twisted pair data bus in MIL-STD 1553 are 31 [13, 17, 18].The entire MIL-STD-1553 physical network infrastructure [3] is depicted in Fig. 6.2 and the MIL-STD-1553 bus requirements are summarized in Table 6.1 [17].

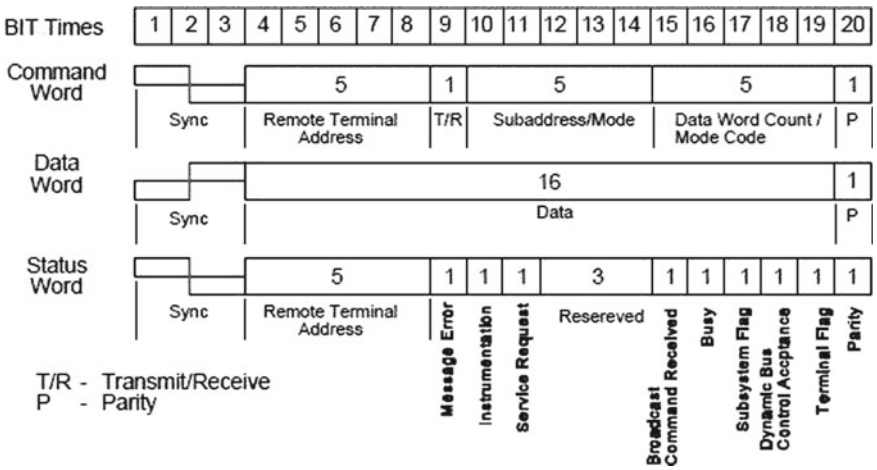


Fig. 6.3 MIL-STD-1553 bus word format structure

6.3.2 MIL-STD-1553 Protocol Format

The sharing of information between devices connected to MIL-STD-1553B data bus is called messages in 1553 protocol. 1553 messages are transmitted in three formats namely, Command Word (CW), Status Word (SW), and pure Data Word (DW) [1, 7, 12]. The detailed word format structure is summarized for more illustration in Fig. 6.3. CW and SW are control words controlling the data transfer mechanism and DW is information word that should be shared among systems connected to data bus [1]. The data encoding and decoding scheme are Manchester bi-phase allowing transition at each bit level, making error detection easy for reliable protocol communication [13]. Due to lower error rate of one-word fault per 10 million words making MIL-STD-1553 highly reliable protocol at an operational speed of 1 Mbps [8, 10]. Table 6.2 summarizes the MIL-STD-1553 network protocol characteristics.

6.3.3 Manchester Encoder/Decoder

MIL-STD-1553 uses Manchester bi-phase encoding scheme to encode the information transmitted through data bus. The scheme is BW expensive, and yet MIL-STD-1553 Manchester bi-phase encoder and decoder are inevitable part of military, avionics, and space applications providing the following key benefits [1] as:

- The timing and clock information can be easily extracted and recovered from encoded data.
- Encoding scheme is immune to channel noise and inter-symbol interference (ISI).

Table 6.2 MIL-STD-1553B network protocol characteristics

Item	Description
Data rate	1 Mbps
Data bits/word	16 bits
Message length	Max 32 data words
Transmission technique	Half duplex
Operation	Asynchronous
Protocol	Command/response
Fault tolerance	Typical dual redundant second redundant in the hot backup status
Terminal types	Remote terminal (RT) Bus controller (BC) Bus monitor (BM)
Types of message transfer	BC to RT transfer RT to BC transfer RT to RT transfer BC to RT (broadcast) RT to RT (broadcast) System control
Number of remote terminals	Max 31 RT
Data words/message	32 data words
Data word format	3 bits sync pattern 16 bits information 1 parity bit
Cable characteristics impedance	70–85 Ω

- Easy error detection capability of order 10^{-12}, even a single bit error can be detected.
- No DC component.

6.3.4 Quality Control Process

According to [19], the introduction of efficient Quality Control Process (QCP) in avionics and automation industry is an important factor in equipment design and the best solution for controlling equipment reliability. The four steps QCP of MIL-STD-1553 architectural attribute quality control evaluation and assessment are as follows:

- Attribute description
- Measurement methodology
- Experimental observation, data collection and analysis
- Scenario-based verification and validation.

Table 6.3 MIL-STD-1553 network quality attributes

Quality domain	Quality attribute	Sub-attributes
Performance	– Data integrity – Information reliability – Communication protocol overhead – System portability – Data bus scalability – Supportability – Information availability – System delay time	– Error detection – Data priority control – Network latency – Throughput – Time synchronization – Timeliness and data accuracy – Cost
Security	– Communication protocol overhead – Throughput	– Network latency – Cost
Bandwidth	– Communication protocol overhead – Throughput	– Network latency – Cost
Architecture	– Data bus device interoperability – System Portability – Data bus access controllability – Data bus scalability – Data bus maintainability – Supportability	– Time synchronization – Timeliness and data accuracy – Cost

List of MIL-STD-1553 quality attributes is categorized in tabular form in Table 6.3.

The next-generation avionics and flight control systems are expected to be highly Software/Hardware collaborative, operating at 200+ Mbps data rate. These systems need to be efficient, cost-effective, reliable, and flexible in harsh environment. MIL-STD-1553 is going to propose design variables must be considered during development process as:

(1) Bus length
(2) Number of stubs
(3) Length of stubs
(4) Location of stubs
(5) Number of remote terminals
(6) Bus encoding scheme

6.4 Comparative Analysis of High-Speed Data Bus Technologies

6.4.1 Traditional MIL-STD-1553 Architecture

MIL-STD-1553 is still a well-suited time division multiplex digital interface for most of the avionics and flight control applications, but each networking technology is selectively decided for certain domain of avionics, not a single technology is considered "best" for all applications [11]. The standard 1553 operates at 1 Mbps still usable for most applications, but emerging high-speed avionics application requires more than 1 Mbps data rate. Once the high-speed applications are in field operation, the data reliability, error control, efficiency, and flexibility comes up in consideration to support the legacy standard, to deliver consistent optimal solution. Two revolutionary approaches are investigated and comparatively analyzed to support the legacy standard (MIL-STD-1553) regarding increased BW efficiency, reliable high-speed data communication, interface flexibility, and performance [3, 5]. In some applications, optimal trade-offs between multiple technologies are preferred. MIL-STD-1553 protocol architectural layers are summarized in Table 6.4.

6.4.2 HyPer-1553TM Data Bus Technology

6.4.2.1 Overview

HyPer-1553TM data bus technology has been implemented by Device Data Corporation (DDC) transmitting data at much higher data rate (>1 Mbps), over the existing MIL-STD-1553 bus by setting up two goals [3], namely:

- HyPer-1553^{TM} technology enables high-speed communication over the existing MIL-TD-1553 cable.
- HyPer-1553^{TM} technology enables high-speed communication at 5 Mbps over the existing MIL-STD-1553 cable and peacefully coexists with legacy 1 Mbps bus cable with no interference.

This new technology helps increase BW between subsystems heavily used in network-centric operations and sensor fusion applications.

Data Device Corporation (DDC) [3, 5, 11, 17, 19] has conducted a series of research studies in real time to implement emerging high-speed data rate (BW expensive) avionics flight control applications, in which the legacy MIL-STD-1553 standard of 1 Mpbs can peacefully co-exist up to 200 Mbps data rate, and yet the standard has a high degree of reliability to scale up to 200+ Mbps in the near future. HyPer-1553TM technology combine's multi-drop bus topology being used for middle-speed networks (from 10 Mbps up to 100 Mbps). Multi-drop bus is cost effective, low-speed network topology using Frequency Division Multiplexing (FDM) to allow concurrent

Table 6.4 Traditional MIL-STD-1553 protocol architectural features

Architectural layer	Architectural features
Physical layer	• Transformer coupling This architectural feature of MIL-STD-1553 serves two purposes: 1. Galvanic isolation 2. Impedance matching Transformer coupling benefit the aircraft flight control system from electromagnetic interference and lightening in harsh environment. Transformer coupled bus minimizes waveform reflections and signal attenuation • Multi-drop linear bus topology The aim of multi-drop linear bus topology tends to lower cost, lower complexity and lower weight • Bus coupler This architectural feature reduces waveform reflections resulting robust physical layer • Fault isolation Fault isolation through series of resistors
Protocol layer	• Highly balanced command/response and deterministic interface • Real time control functions and error detection capability • 1553 periodic data transfer bus • MIL-STD-1553 increase transmission accuracy • MIL-STD-1553 protocol gets lower delay and jitter
Reliable data link layer	• MIL-STD-1553 command/response acknowledgment facility for retransmission combines with error detection • Small payload size (64 bytes) • Dual redundant multi-drop bus topology has been used for systems requiring high reliability • Time synchronization in MIL-STD-1553 enable reliable data transfer

low-speed and high-speed data communication over MIL-STD-1553 cable. Multi-drop bus eliminates active hubs and switches significantly reducing size, weight, power, and cost. HyPer-1553TM technology can provide solutions for efficient and reliable high-speed communication on multi-drop bus using advanced signaling and filtering techniques. The high-speed data rate communication depends heavily on the length of bus, number of stubs, and stub length. HyPer-1553TM technology is scalable and flexible approach by adding high data rate applications, sharing the same cable without interfering with the legacy 1 Mbps interface.

6.4.2.2 Test Result Analysis

DDC has conducted 2 h onboard flight demonstration, testing HyPer-1553^{TM} digital bus using USAF F-15 E1 strike eagle fighter in December 2005 [3]. The imagery data was transmitted between rugged computer in the forward avionics bay and weapon mounted on wing pylon station. During the test, the team was successful to transmit

the data at a rate of 40 Mbps over legacy MIL-STD-1553 bus in parallel with MIL-STD-1553 1 Mbps for 2 h. The team also transferred data at rate of 80 and 120 Mbps on second 1553 bus dedicated to high speed. The received high-speed imagery data was error free validating the HyPer-1553^{TM} Multi-drop bus reliability, efficiency, and flexibility over legacy MIL-STD-1553 1 Mbps cable.

6.4.3 Turbo 1553 Approach

6.4.3.1 Overview

MIL-STD-1553 has a well-established set of design guidelines for a network operating at 1 Mbps. In addition to over 30 years of in-service history, there is a strong analytical foundation for these guidelines which is well documented in MIL-HDBK-1553A. The key design variables in a 1553 network are bus length, number of stubs, location of stubs, and length of the stubs. The concepts defined in the standard and the handbook can be extended to data rates above 1 Mbps. The question becomes what impact would a higher data rate have on these design variables and the resulting performance of the network.

The first step toward an implementation of Turbo 1553 is to understand the impact of higher frequency on attenuation and phase distortion. Attenuation impacts the amplitude of the signal that is presented to the receiver, and as such impacts the resulting Signal-to-Noise Ratio (SNR). SNR is a key benchmark in defining the throughput capacity and Bit Error Rate (BER) of a network. Phase distortion, also referred to as jitter or zero crossing error, impacts the relative timing of pulses which in turn can lead to problems with inter-symbol interference which also has an impact on the bit error rate of the receiver. MIL-STD-1553 test network settings for Turbo-1553 (Scenario 1, Scenario 2) are summarized in Table 6.5.

6.4.3.2 Test Result Analysis

Turbo-1553 test network experimental setup is illustrated in Fig. 6.4 to evaluate MIL-STD-1553 bus interface running at higher speed (>1 Mbps). The length of the bus is 460 feet with 10 interconnecting stubs ranging in length from 1 foot to 5 feet with terminal impedance of 78 Ω is given for illustration. Communication was tested between 1553 bus controller (1553-BC-Terminal 1) and 1553 Remote Terminals (1553-RT-2, 1553 RT-3, and 1553 RT-4). Two test network scenarios, one describing Turbo-1553 with 10 interconnecting stubs and second with no stubs are tested and compared. Two Turbo-1553 network scenarios are tested and compared in terms of generated frequency response and corresponding phase distortion running at 5 Mbps.

Frequency response of 460 feet cable with 10 interconnecting stubs versus 460 feet cable with no stubs is measured as illustrated in Fig. 6.5. From Fig. 6.5, it is clear

Table 6.5 Turbo-1553 test network settings

Test settings	Turbo-1553 scenario 1	Turbo-1553 scenario 2	HyPer-1553
Data rate	5 Mbps	5 Mbps	100+ Mbps
Bus length	460 feet	460 feet	
Stub length	1–5 feet	1–5 feet	
Number of interconnected stubs	10 stubs	0 stubs	
BC	Terminal-1	Terminal-1	
RT	RT-2, RT-3, RT-4	RT-2, RT-3, RT-4	
Signal loss	12.6 dB	12.6 dB	
Stub voltage	1.4v P-P	1.4v P-P	
Transmitter voltage	6 V minimum	6 V minimum	
Attenuation	RT-2: least RT-3: moderate RT-4: most		
Phase distortion	RT-2: largest due to reflection RT-3: moderate due to reflection RT-4: largest due to dispersion		
MIL-STD-1553 terminator	78 Ω		

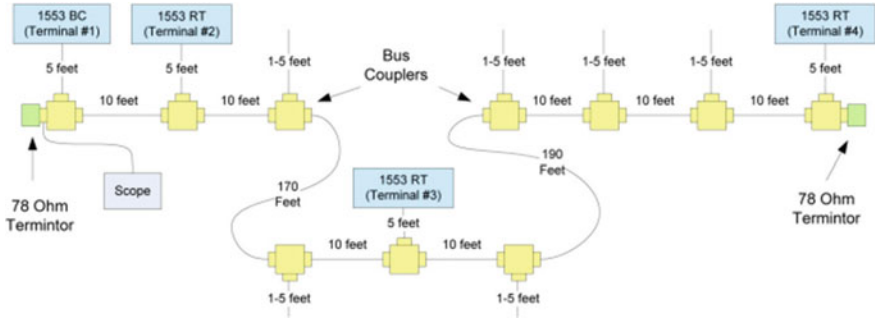


Fig. 6.4 Turbo-1553 network experimental setup

that at higher frequencies, the signal attenuation will be more in cable with 10 stubs as compared to the one having no stub.

The similar test experiment has been conducted by DDC using 300 feet cable, and corresponding frequency response ranging from 300 kHz to 10 MHz is generated as shown in Fig. 6.6. The test result shows that the observed signal attenuation of −2 dB at 1 MHz and −5 dB at 5 MHz, both are within the specified MIL-STD-1553 range of 12.6 dB. The signal attenuation and phase distortion measured from three remote terminals are well within the specified receiver range. The signal attenuation and

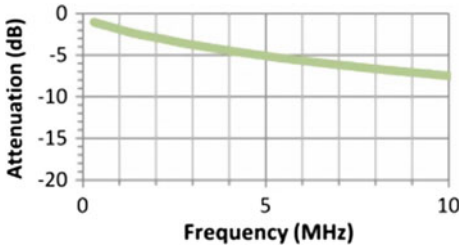


Fig. 6.5 Frequency response of 300 feet of 1553 cable

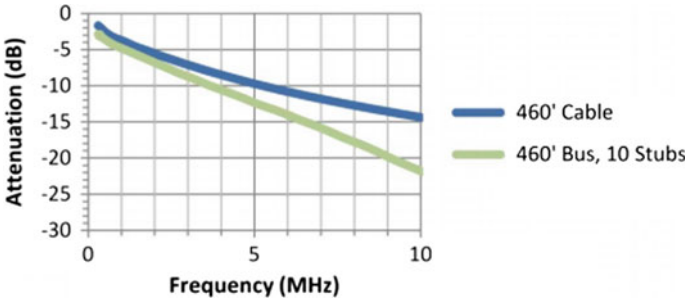


Fig. 6.6 Insertion loss of 460 feet cable and 460 feet bus from 300 kHz to 20.3 MHz

phase distortion are also affected by the number of stubs connected to bus and stub length. The test result shows that bus controller was communicating with Remote Terminals efficiently and reliably in high-speed applications.

6.4.4 Tools for Testing and Simulation

Design testing and simulation is the most important phase of Product Development Life Cycle (PDLC) and up to 80% of the total aircraft cost is set during the early phase of development [16, 19]. The goal of MIL-STD-1553 bus testing and simulation is to identify and mitigate the risk and architectural defects in order to ensure the product reliability and quality assurance. The architectural defect/bugs removal process is highly expensive once the product is launched for an operation in real time. So, continuous integration testing and simulation during development phase are highly preferable to control the equipment reliability [19]. MIL-STD-1553 testing levels are described in [16], and testing level classifications are summarized in Table 6.6 with test description and type of scenario.

Table 6.6 MIL-STD-1553 levels of testing

Testing level	Description	Scenario
Field/operational level testing (high level testing)	Real time on board test to verify and validate MIL-STD-1553 digital interface installed	HyPer-1553 bus testing (at 4 Mbps) 2 h onboard flight USAF F15-E1 strike eagle fighter in December 2005 [3]
System integration testing	Integrating MIL-STD-1553 system SW/HW components to test bus interoperability, scalability and flexibility to ensure the overall system reliability [20]	System integration testing of KAI ETS system model including EVA card MIL-STD-1553B card mission computer [20]
Production testing	Testing unit functions/component/operation/unit level fault assurance [12]	Xilinx Spartan FPGA kit using Xilinx ISA [13]
Design verification	Preproduction phase testing model specification requirement testing	ALTARICA [19] functional and dysfunctional behavior for component reliability
Developmental testing	Circuit level testing, low-level design	ModelSim [13, 21] Verilog HDL [1, 13, 15, 21] VHDL [13, 21, 22] OPNET and NS-2 [23]

6.5 Conclusions and Future Works

In this chapter, MIL-STD-1553 digital time division multiplex serial interface bus is investigated for reliability and flexibility in high data rate applications. The standard has 40+ years of quality-controlled flight experience, lending itself as a proven technology widely used in flight control and avionics applications. MIL-STD-1553 is an inexpensive technology by dramatically reducing the size, weight, power, and cost in the last 10 years [19]. The traditional MIL-STD-1553 has been investigated and analyzed for high data rate applications using 1553 derivative technologies such as Turbo-1553 (50–100+ Mbps) and HyPer-1553TM (5 Mbps). The resulting test network analysis verifies that at higher data rate, the standard provides reliable data transfer (error free) and flexible control over the existing legacy 1553 standard. The next-generation avionics, space, automation, and digital control systems needs to investigate optimal solution for high-speed data bus with high degree of compatibility to support MIL-STD-1553 [24, 25].

The next-generation avionics and flight control systems are predicted to support large data files transfer, handling volumetric images, and videos transmission in critical flight mission and finally providing abstract level Software/Hardware collaboration. The emphasis would be on high degree of digital data bus reliability, intelligent error detection, bus access control, flexibility, and component supportability. Next-generation Manchester bi-phase encoder and decoder design for high-speed data communication would be preferable approach to improve the overall system reliability.

References

1. Jose J (2013) Design of Manchester II bi-phase encoder for MIL-STD-1553 protocol. In: International multi-conference on automation, computing, communication, control and compressed sensing, pp 240–245, 22–23 Mar 2013
2. Lundy GM, Christensen PH (1990) Specification of the MIL-standard 1553 protocol using systems of communicating machines. In: Conference record on IEEE military communications conference, vol 3, pp 1049–1053, 30 Sept–3 Oct 1990
3. Hegarty MG (2010) MIL-STD-1553 evolves with the times. In: Data device corporation white paper, June 2010
4. Hegarty GM (2004) High performance 1553: a feasibility study. In: The 23rd digital avionics systems conference, vol 2, pp 1–9, 24–28 Oct 2004
5. Glass M (2007) Buses and networks for contemporary avionics. In: Data device corporation white paper, Nov 2007
6. Mock M, Schemmer S, Nett E (2000) Evaluating a wireless real-time communication protocol on Windows NT and WaveLAN. In: IEEE international workshop on factory communication systems, pp 247–254
7. Zhang YX, Wei-Gong Z, Quan Z, Rui D, Yuan-Yuan S (2010) The design of 1553B communication bus based of BU-61580. In: The 5th IEEE conference on industrial electronics and applications, pp 1920–1923, 15–17 June 2010
8. Truitt RB, Sanchez E, Garis M (2004) Using open networking standards over MIL-STD-1553 networks. In: AUTOTESTCON Proceedings, pp 117–123, 20–23 Sept 2004
9. Truitt RB, Sanchez E, Garis M (2005) Using open networking standards over MIL-STD-1553 networks. IEEE Aerosp Electron Syst Mag 20(3):29–34
10. Li Z, Junshe A (2013) Study on the low power technologies of 1553B bus. In: IEEE international conference on signal processing, communication and computing, pp 1–5, 5–8 Aug 2013
11. Hegarty MG (2005) Avionics network technology. In: Data device cooperation white paper
12. Schuh RA (1988) An overview of the 1553 bus with testing and simulation considerations. In: 5th IEEE instrumentation and measurement technology conference, pp 20–25, 20–22 April 1988
13. Jose J, Varghese S (2012) Design of 1553 protocol controller for reliable data transfer in aircrafts. In: 12th international conference on intelligent systems design and applications, pp 686–691, 27–29 Nov 2012
14. Diao LF, Dai M, Lei Jian JM (2010) Application of IP core technology to the 1553B bus data traffic. In: Asia pacific conference on postgraduate research in microelectronics and electronics, pp 338–342, 22–24 Sept 2010
15. Chunping H, Shuai W, Qing W, Hao Z (2013) Performance analysis of high-speed MIL-STD-1553 bus system using DMT technology. In: 8th international conference on computer science & education, pp 533–536, 26–28 April 2013
16. Furgerson J (2010) MIL-STD-1553 tutorial. AIM GmbH, Nov 2010 (online). Available: http://www.aim-online.com/pdf/OVW1553.PDF
17. Data Device Corporation (1998) MIL-STD-1553 designer's guide 6th edn
18. MIL-STD-1553B (1978) Aircraft internal time-division multiplexing data bus, Department of Defense, Washington, D.C. (online). Available: http://www.ballardtech.com/T/MIL-STD-1553BNII.PDF
19. Hegarty MG (2010) A practical approach to commercial aircraft data buses. In: Data device corporation white paper, Mar 2010
20. Kun SY, Sang WY, Chae II S (2009) New architecture for improving performance in embedded training system using embedded virtual avionics. In: IEEE/AIAA 28th digital avionics systems conference, pp 1–6, 23–29 Oct 2009
21. Kai H, Kai L, Cheng L (2012) The design and implementation of an extended 1553B bus IP core to support large file transfer. In: 7th international conference on computing and convergence technology, pp 494–498, 3–5 Dec 2012

22. Wei H, Xiaojuan L, Yong G, Zhiping S, Lingling D, Jie Z (2012) Formal verification for space wire communication protocol based on environment state machine. In: 8th international conference on wireless communications, networking and mobile computing, pp 1–4, 21–23 Sept 2012
23. Hendricks S, Duren R (2000) Using OPNET to evaluate fiber channel as an avionics interconnection system. In: Digital avionics systems conference
24. Gorringe C (2013) Bus testing in a modern era. In: IEEE AUTOTESTCON, pp 1–10, 16–19 Sept 2013
25. Jifeng L, Minggang C (2011) Design of 1553B avionics bus interface chip based on FPGA. In: International conference on electronics, communications and control, pp 3642–3645, 9–11 Sept 2011

Chapter 7
Research and Design of 1553B Protocol Bus Control Unit

7.1 Introduction

The digital data bus MIL-STD-1553B was designed in the early 1970s to replace analog point-to-point wire bundles between electronic instrumentation. The latest version of the serial Local Area Network (LAN) for military avionics known as MIL-STD-1553B was issued in 1978. After 30 years of familiarity and reliable products, the data bus continues to be the most popular militarized network.

In avionics system, the 1553B interface board is an important part in the whole system and it mainly integrates data in bus, share resources, coordination tasks, and fault-tolerant reconstruction [1]. The technology of compatible with high-performance general-purpose microcomputer and large-scale integrated circuits is widely applied to complete interface communication from the 80s in our country. The bus controller not only insures sending or receiving commands is correct, but also monitors the bus status. Now, we must to adopt an expensive foreign 1553B protocol processor in order to design BCU, the chapter introduces a new design method of 1553B protocol BCU (Bus Control Unit), the bus controller is an important component part of the communication system, and the main hardware platform is the FPGA (Field-Programmable Gate Array) chip called XC2V2000 of Xilinx company. Based on in-depth study, 1553B bus transport protocols and foreign design method of chip, and combined with popular EDA technology, the chapter designed successfully the digital 1553B MIL-STD-BCU under the top-down design method, and proved to be correct on self-designed experiment board also [2].

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_7

7.2 1553B Protocol

7.2.1 Hardware Characteristics

The MIL-STD-1553B bus has four main elements:

- A bus controller that manages the information flow.
- Remote terminals that interface one or more simple subsystems to the data bus and respond to commands from the bus controller.
- The bus monitor that is used for data bus testing.
- Data bus components (bus couplers, cabling, terminators, and connectors). Data is sequentially transmitted and received in a multiplexing scheme over two copper wires from computer to computer at a rate of 1 megabit per second.

7.2.2 Encoding

The data encode shall be Manchester II bi-phase level. A logic one shall be transmitted as a bipolar coded signal 1/0 (i.e., a positive pulse followed by a negative pulse). A logic zero shall be a bipolar coded signal 0/1 (i.e., a negative pulse followed by a positive pulse). A transition through zero occurs at the midpoint of each bit time. The transmission bit rate on the bus shall be 1.0 megabit per second. The command sync waveform shall be an invalid Manchester waveform. The width shall be three-bit times, with the sync waveform being positive for the first one and one-half bit times, and then negative for the following one and one-half bit times. If the next bit following the sync waveform is a logic zero, then the last half of the sync waveform will have an apparent width of two clock periods due to the Manchester encoding.

7.2.3 Word and Message

The word formats include the command, data, and status words. The word size shall be 16 bits plus the sync waveform and the parity bit for a total of 20 bits times. A command word shall be comprised of a sync waveform, remote terminal address field, transmit/receive (TIR) bit, sub-address/mode field, word count/mode code field, and a Parity (P) bit. A data word shall be comprised of a sync waveform, data bits, and a parity bit. A status word shall be comprised of a sync waveform, RT address, message error bit, instrumentation bit, service request bit, three reserved bits, broadcast command received bit, busy bit, subsystem flag bit, dynamic bus control acceptance bit, terminal flag bit, and a parity bit. For optional broadcast operation, transmission of the status word shall be suppressed. The messages transmitted on the data bus includes bus controller to remote terminal transfers, remote terminal to bus

controller transfers, remote terminal to remote terminal transfers, mode command without data word, mode command with data word (transmit), mode command with data word (receive), and optional broadcast command [3].

7.2.4 Hierarchical Division

The communication system consists of physical layer (PHY), Data Link Layer (DLL), transport layer, and application layer.

- PHY: Provide communications medium, the upper management of physical media stream transmission, to ensure the required transmission characteristics with 1553B information bits sent to the data link layer.
- DLL: In order to transmit reliable, the layer defines the data order in accordance with 1553B protocol, and detects communication errors and reports to transport layer on time.
- Transport Layer: Query and transmit nodes messages, handle errors, and switch channels.
- Application Layer: Services facilitate communication between software applications and lower layer network services so that the network can interpret an application's request and, in turn, the application can interpret data sent from the network [4–8].

7.3 BCU Design

BCU is the only bus device in the transmission system, responsible for scheduling the data in the bus, and sending control commands to other terminals [9]. BCU functions is as follows:

- Receiving Data Words: The function includes simulating receiver, synchronous detection, makes conversion between parallel data and serial data, Manchester code detection, parity, bit/word count, and so on.
- Sending Data Words: The function includes generating CLK and sync waveform, encoding, controlling sending, and so on.
- Handling Words/Messages: The function includes decoding command words, receiving and decoding status words, producing interruptions, and detecting errors.

BC Logical structure based on FPGA is shown in Fig. 7.1 and BC workflow is shown in Fig. 7.2.

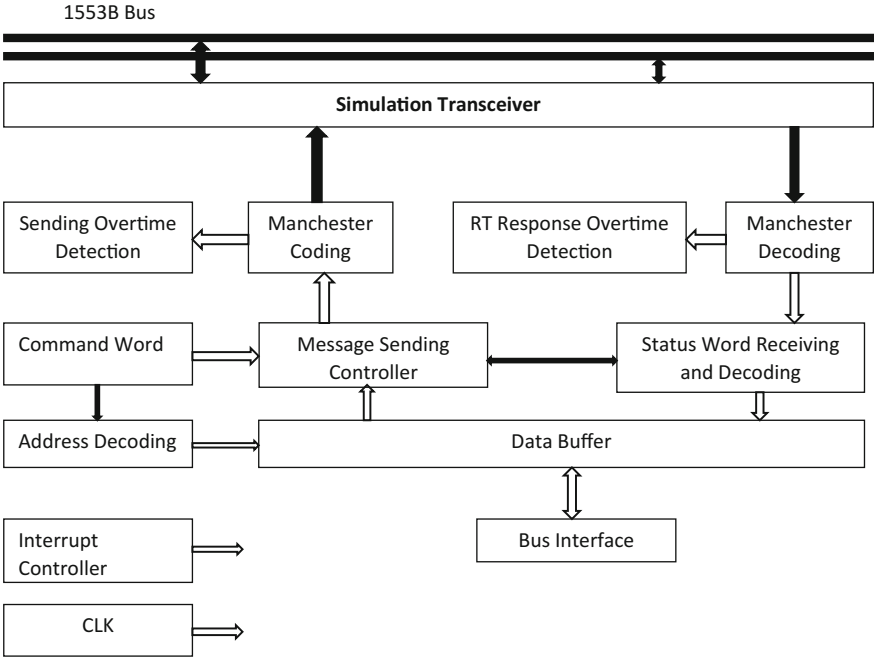
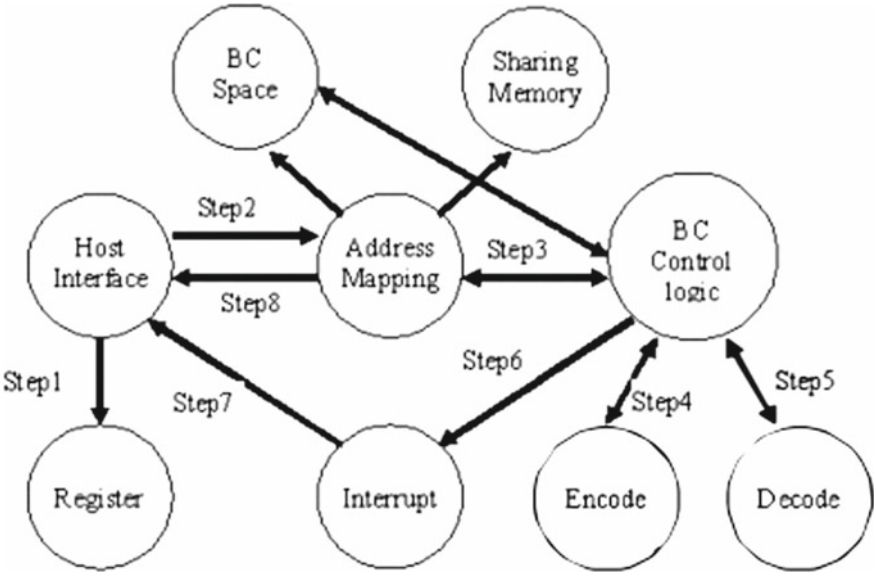


Fig. 7.1 BC logical structure on FPGA

7.3.1 Decoding Unit

When the bus interface board receiving serial data, FPGA getting the data by simulation transceiver also, and searching sync waveform at the same time, if success, it begins to detect Manchester II encode, parity, and bit/word count, etc., finally, the handled 16-bit parallel data can be decoded by post-processing modules [8].

- Sync waveform detection of command words and state words: With input frequency of 12 MHz, after resetting state machine, the system prepares to detect sync waveform, and when the bus data state from inactive turned into active, it starts to detect sync waveform of command words and status words. The 1553B protocol is specified as sync waveform is composed of 1.5 μs by the high level and 1.5 μs low level, and also defines there has a new command word or a state word when sampled continuous 18 high level and 18 low levels. Because of all kinds of interferences in bus, low level cannot be turned into high level in time, often take some time. We can't get the waveform with ideal steep rise along too, so both the number of high level and low level shouldn't be 18, but 14 and 16 in the chapter.
- Decode: As shown in Fig. 7.3, the 12 MHz frequency is split six equal copies first, and sample with the 2 MHz frequency. We will get two sample data at one quarter and three-quarters position of each cycles, then judge whether the code is



Instruction
Step l: Set Register of BC Work Pattern.
Step 2: Set Parameters of BC and Write Sync/Async messages.
Step 3: Handle Sync/A sync messages in BC.
Step 4: Send command words and data words.
Step 5: Receive Status words and data words.
Step 6: Processing Procedure of Messages is over.
Step 7: Presents an interruption.
Step 8: Get Handled Results by Host.

Fig. 7.2 BC workflow

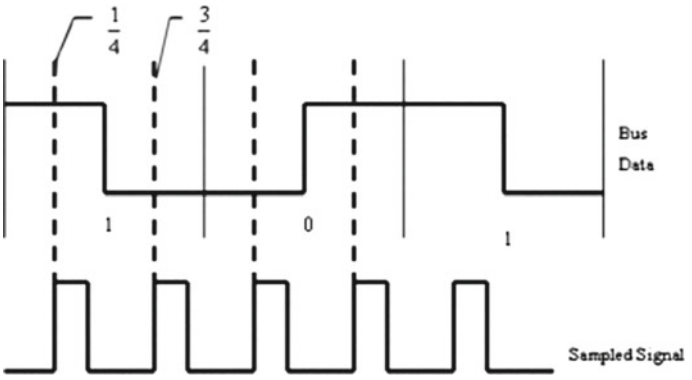


Fig. 7.3 Data sampling by sampled signal

in correct accord with the data, if two values are different, shows the bus data is correct.

- Parity: The last bit in the word shall be used for parity over the preceding 16 bits. Odd parity shall be utilized. If it is not correct, should be set PARITY_ERR = l, and send the message to other relevant modules.
- Sync waveform detection of data words: The data sync waveform shall be an invalid Manchester waveform. The width shall be three-bit times, with the waveform being negative for the first one and one-half bit times, and then positive for the following one and one-half bit times. Note that if the bits preceding and following the sync are logic ones, then the apparent width of the sync waveform will be increased to four-bit times.

7.3.2 Data Encode Unit

The data encode shall be Manchester II bi-phase level. A logic one shall be transmitted as a bipolar coded signal 1/0 (i.e., a positive pulse followed by a negative pulse). A logic zero shall be a bipolar coded signal 0/1 (i.e., a negative pulse followed by a positive pulse). A transition through zero occurs at the midpoint of each bit time. The unit function includes non-RZ (NRZ) code tum into Manchester code, sync waveform encode, make a parity bit, and parallel/serial conversion.

The encode process is as follows: (1) make a sync waveform, (2) parallel/serial conversion, (3) make a parity bit, and (4) encode the 16 data bits and a parity bit. With the input frequency of 12 MHz, when TX-CSW is true, place a sync waveform bits in front of valid data bits in the command word, and encode by Manchester encoding.

7.3.3 Command Words Decode Unit

As one part of decoding unit, its main function is to decode sent command words and received returned status words by RT in BC mode, and send control commands to other units. The unit workflow includes set MODE = “00” (“00”:BC mode, “01”:RT, “10”:MT) before send message, compute RT address, sub-address, message type, send word count, address in decoder, the initial address, and address selection signal in light of 1553B protocols. In addition to, the next step is necessary also, that getting control commands accord with read data from RAM or wrote RAM.

7.3.4 Send Control Unit

The main function is to get data for sending of encoder, determine the start-time for sending data to RT, and make sync waveform. The unit workflow includes load command words for encoding to encoder, output synchronizing single at the same

time, and judge whether there are data words for sending behind command words, if so, beginning to encode and send the data, otherwise, the work status turns into idle.

7.3.5 Status Words Receive Control and Decode Unit

A status word shall be comprised of a sync waveform, RT address, message error bit, instrumentation bit, service request bit, three reserved bits, broadcast command received bit, busy bit, subsystem flag bit, dynamic bus control acceptance bit, terminal flag bit, and a parity bit. The main function is to determine whether continue receive bus data according to the output parameters that are set by operational mode and input signals. Through decoding status words that RT returned, we may know the system working status in BC mode.

The unit workflow as follows, first, determined whether the data is status words according to sync waveform flag of decode unit. Second, it needs to determine whether the RT address is correct according to the first five bits of a status word, if it is not, should be set RTADDR-ERR = 1, otherwise, decode status words. Finally, it will receive data words behind of status words under the control of command words.

7.3.6 Address Decode Unit

The main function is to decode address in command words decode unit, load data, initial address, initial data words counter, and initial RX-RDY, TX-RDY. It is noteworthy that the initial RAM address as input is computed by RT address list in command words, where there is an RT address, there is a buffer in RAM, and the value equal to the start position of allocating buffer in RAM.

Under the control of the address signal, initialize address first, then receive or send data words and plus one to the address according to feedback of signals come from the encode/decode unit, until all data have received or sent. The operation will be triggered by the data ready signal of RX-RDY and TX-RDY. Because the time of generating the address signals is later than the ready signals, so the operation is only triggered after a certain delay.

7.3.7 Send Overtime Detection Unit

We can grasp the time of send data easily, according to the interval between send and receive message. There is a counter for monitoring the output of diver, if the value is greater than 800 μs when sending data, it will stop to send and encode, generate overtime signals for resetting finally.

7.3.8 Error Detection Unit

In order to ensure the communication system is reliable, we will adopt the Automatic Repeat Request (ARQ) protocols. The chapter introduces a new detection method for finding errors, named as software design approach. The main function is to check encode/decode data adopting the technology, determine the RT response overtime, detect word count, and generate errors interrupts.

- Word Count Detection: The main function is to check the number of sent data words, the number of received data words, and control the count of words by the positive and negative value of Manchester.
- RT Response Overtime: The main function is to determine whether status words are received in time after send command words or data words. At the same time, it also determines the time of RT return status words.
- Generate Interrupts: Set INT $1 = 1$ accord to error types such as parity error, counter error, RT response overtime, RT address error, etc.

7.3.9 DSP Communication Interface

The chapter introduces a main processor named as ADSP-21161. By using the parallel communication interface, we completed information exchange and data storage between the main processor and the protocol chip on the address bus, the data bus and the control bus.

- Parallel Communication: The main function is to control the communication between protocol modules and DSP, it includes data signal, address signal, WRIRD, and MSO-MS3 of DSP.
- Storage Interface: It provide interfaces between reading/writing signal in the chip and visiting signal in DSP, and avoid access conflict on reading or writing at the same time under the control of priority.
- Dual Port RAM: It provides a buffer in the chip, and adopts the $2K \times 16$ RAM in the chapter.

7.4 Logic Emulation

The chapter introduces the simulation process and analysis of simulation result about reading or writing to CPU, state transition of controller and protocol state machine. Figure 7.4 shows sending command words to RT and receiving returned status words of RT.

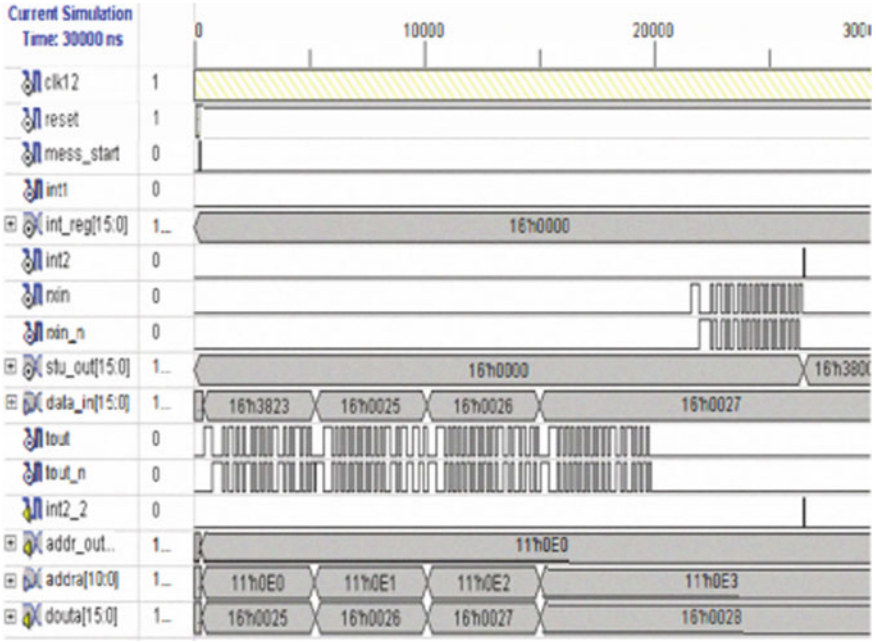


Fig. 7.4 The simulation result of sending a command word

7.5 Conclusions

In this chapter, the 1553B protocol BCU is designed successfully based on the deep research for 1553B protocol. The design method shows both a good grasp of design planning and focused on technical details. The results of the simulation show the effectiveness of this method.

References

1. Yang J (1995) The system of modem military aviation electronic communications. Aeronaut Comput Tech, 11–15
2. Pizzica S, Raytheon Company (2001) Open systems architecture solutions for military avionics testing. IEEE AESS Syst Mag, 45–47
3. No. 301 Institute of Aeronautics and Astronautics, Application note for MIL-HDBK-1553 multiple data bus
4. Department of defense interface standard for digital time division command/response multiplex data bus (l975)
5. Wang Z (2006) Research of bus communications in integrated avionics system based on MIL-STD-1553B. NanJing University of Science and Technology
6. Murdork JR, Koenig JR (2001) Open systems avionics network to replace MIL-STD-1553. IEEE AESS Syst Mag, 34–36

7. DDC Corporation (1999) MIL-STD-1553 designer's guide
8. Ji Jao S (2004) Design of 1553B decoder based on FPGA. In: Microcontroller & embedded system, pp 42–44
9. Bai Y, Zhou Z, Chen J (1996) The implementation of MIL-STD-1553B processor. In: ASIC 2nd international conference, pp 8/1–8/7, 21–24 Oct 1996

Part II
Industrial Wireless Sensor Networks

Part II of the book titled Industrial Wireless Sensor Networks includes 11 research proposals analyzing and evaluating Industrial Wireless Sensor Networks (IWSN) in terms of wireless technology. Such aspect is highlighted by key points composed of Medium Access Control (MAC) mechanisms, wireless communication standards for industrial field. In additions, applications of such networks from environmental sensing, condition monitoring, and process automation applications are specified. Designing appropriate networks is based on the specific requirements of applications. It points out the technological challenges of deploying WSNs in the industrial environment as well as proposed solutions to the issues. An extensive list of IWSN commercial solutions and service providers are provided, and future trends in the field of IWSNs are summarized.

Chapter 8
An Overview on Wireless Sensor Networks

8.1 Introduction

A wireless sensor network (WSN) can be generally described as a network of nodes that cooperatively sense and may control the environment enabling interaction between persons or computers and the surrounding environment [1]. On one hand, WSNs enable new applications and thus new possible markets; on the other hand, the design is affected by several constraints that call for new paradigms [2, 3]. In fact, the activity of sensing, processing, and communication under limited amount of energy, ignites a cross-layer design approach typically requiring the joint consideration of distributed signal/data processing, medium access control, and communication protocols [4]. WSNs have several common aspects with wireless ad hoc network [5] and in many cases, they are simply considered as a special case of them. This could be lead to erroneous conclusions, especially when protocols and algorithms designed for ad hoc networks are used in WSN. For this reason, in Sect. 8.2, an appropriate definition of WSN and discussion is provided.

In Sect. 8.3, the main application areas for WSNs are categorized according to the type of information measured or carried by the network. Applications, on top of the stack, set requirements that drive the selection of protocols and transmission techniques; at the other end, the wireless channel poses constraints to the communication capabilities and performance. Based on the requirements set by applications and the constraints posed by the wireless channel, the communication protocols and techniques are selected. The main features in WSNs are described in Sect. 8.4. Specifically, the design of energy-efficient communication protocols is a very peculiar issue of WSNs, without significant precedent in wireless network history.

Generally, when a node is in transmitting mode, the transceiver drains much more current from the battery than the microprocessor in active state or the sensors and the memory chip. The ratio between the energy needed for transmitting and for processing a bit of information is usually assumed to be much larger than one (more than 100 or 1000 in most commercial platforms). For this reason, the communication protocols

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_8

need to be designed according to paradigms of energy efficiency, while this constraint is less restrictive for processing tasks. Then, the design of energy-efficient communication protocols is a very peculiar issue of WSNs, without significant precedent in wireless network history. Most of the literature on WSNs deals with the design of energy-efficient protocols, neglecting the role of the energy consumed when processing data inside the node, and conclude that the transceiver is the part responsible for the consumption of most energy. On the other hand, data processing in WSNs may require consuming tasks to be performed at the microprocessor, much longer than the actual length of time a transceiver spends in transmit mode. This can cause a significant energy consumption by the microprocessor, even comparable to the energy consumed during transmission, or reception, by the transceiver. Thus, the general rule that the design of communication protocol design is much more important than that of the processing task scheduling is not always true.

8.2 Wireless Sensor Networks

A WSN can be defined as a network of devices, denoted as nodes, which can sense the environment and communicate the information gathered from the monitored field (e.g., an area or volume) through wireless links [6, 7]. The data is forwarded, possibly via multiple hops, to a sink (sometimes denoted as controller or monitor) that can use it locally or is connected to other networks (e.g., the internet) through a gateway. The nodes can be stationary or moving. They can be aware of their location or not and they can be homogeneous or not.

This is a traditional single-sink WSN (see Fig. 8.1, left part). Almost all scientific papers in the literature deal with such a definition. This single-sink scenario suffers from the lack of scalability: by increasing the number of nodes, the amount of data gathered by the sink increases and once its capacity is reached, the network size cannot be augmented. Moreover, for reasons related to MAC and routing aspects, network performance cannot be considered independent from the network size.

A more general scenario includes multiple sinks in the network (see Fig. 8.1, right part) [8]. Given a level of node density, a larger number of sinks will decrease the probability of isolated clusters of nodes that cannot deliver their data owing to unfortunate signal propagation conditions. In principle, a multiple-sink WSN can be scalable (i.e., the same performance can be achieved even by increasing the number of nodes), while this is clearly not true for a single-sink network [9].

However, a multi-sink WSN does not represent a trivial extension of a single-sink case for the network engineer. In many cases nodes, send the data collected to one of the sinks, selected among many, which forward the data to the gateway, toward the final user (see Fig. 8.1, right part). From the protocol viewpoint, this means that a selection can be done, based on a suitable criteria that could be, for example, minimum delay, maximum throughput, minimum number of hops, etc. Therefore, the presence of multiple sinks ensures better network performance with

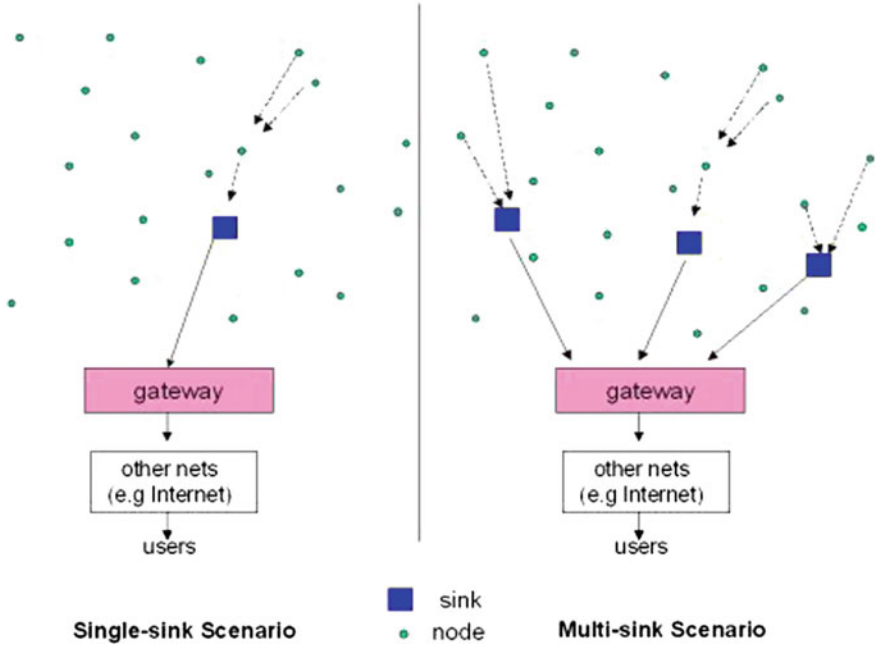


Fig. 8.1 Left part: single-sink WSN. Right part: multi-sink scenario

respect to the single-sink case (assuming the same number of nodes is deployed over the same area), but the communication protocols must be more complex and should be designed according to suitable criteria.

8.3 Network Topologies of Wireless Sensor Networks

To help our discussion about Wireless Sensor Networks, we refer the reader to Fig. 8.2, which depicts a simple star network consisting of six nodes. Using IEEE 802.15.4 terminology, this collection of nodes is termed a PAN; and it is assumed to span a small (G 10 m) geographical area. Additionally, there are two types of nodes defined in the standard; a Full-function device (FFD) and a Reduced function device (RFD). From the PAN control and multiple-access point of view, an FFD contains the software that enables PAN initiation, network formation, and control of the wireless channel for multiple accesses among the RFDs.

An FFD is commonly referred to as a "coordinator" due to its ability to provide the above functions. In the figure, the FFD node is depicted in the center of the PAN while the RFD nodes are shown surrounding the coordinator. The arrows indicate that the RFD devices are logically associated with the coordinator and rely on it for multiple-access services and data transport.

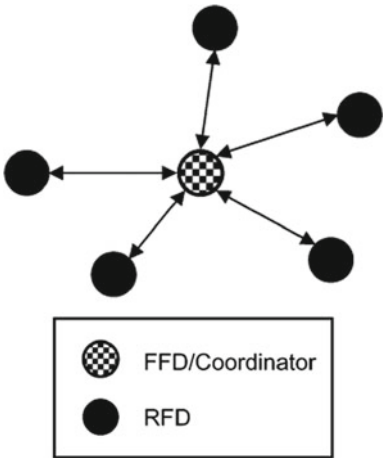


Fig. 8.2 A simple sensor network with a star topology

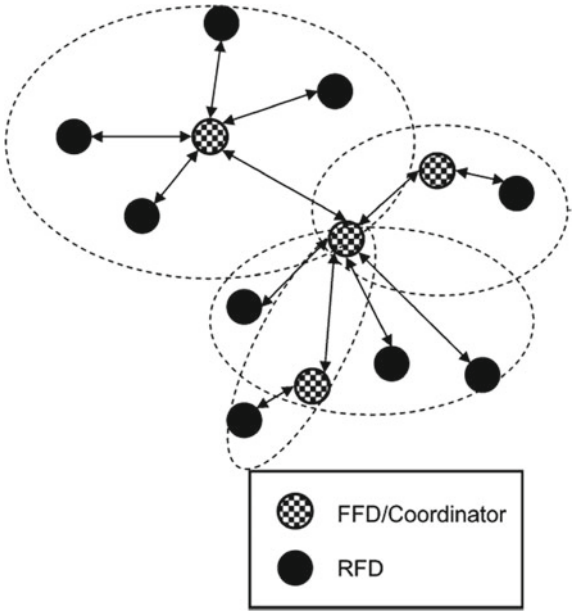


Fig. 8.3 Sensor network with a tree topology

Figure 8.3 shows another example of a sensor network topology, typically referred to as a tree network. In this figure, we again consider both FFD and RFD devices as in Fig. 8.2.

The tree network can be viewed as an amalgamation of star networks (depicted by the dashed circles) where the star networks are connected together by linking the FFDs in each star together. Note here that data may need to be routed through multiple hops if devices want to communicate outside of their local star network.

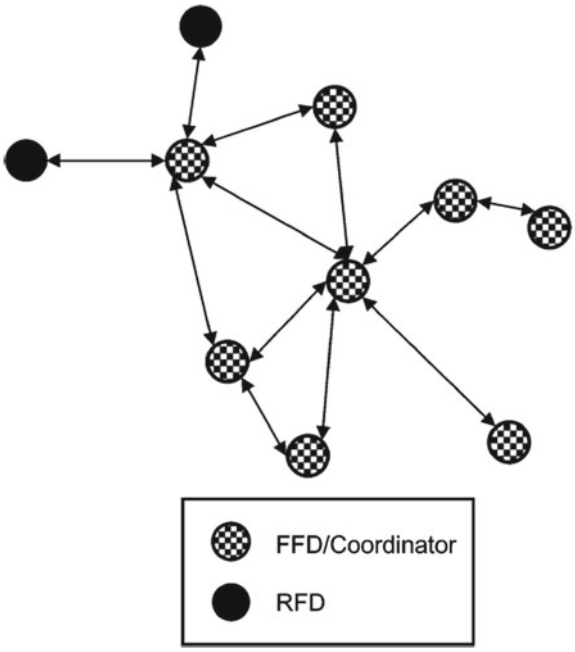


Fig. 8.4 Sensor network with a mesh topology

A third topology to consider is a mesh topology, which is similar to the multi-hop tree topology but with the addition of multiple links among the devices. (In a tree network, there exists only one path between any two devices.) The mesh topology in Fig. 8.4 provides reliability to the network in the form of redundant paths among the devices so, in the event of device or link failure, data may be rerouted.

8.4 Applications of WSNs

The variety of possible applications of WSNs to the real world is practically unlimited, from environmental monitoring, health care, positioning and tracking, to logistic, localization, and so on. A possible classification for applications is provided in this section. It is important to underline that the application strongly affects the choice of the wireless technology to be used. Once application requirements are set, in fact, the designer has to select the technology which allows to satisfy these requirements. To this aim the knowledge of the features, advantages and disadvantages of the different technologies is fundamental.

8.4.1 Application Classification

One of the possible classifications distinguishes applications according to the type of data that must be gathered in the network. Almost any application, in fact, could be classified into two categories: event detection (ED) and Spatial Process Estimation (SPE).

In the first case, sensors are deployed to detect an event, for example, a fire in a forest, a quake, etc. [10]. Signal processing within devices is very simple, owing to the fact that each device has to compare the measured quantity with a given threshold and to send the binary information to the sink(s). The density of nodes must ensure that the event is detected and forwarded to the sink(s) with a suitable probability of success while maintaining a low probability of false alarm. The detection of the Phenomenon of Interest (POI) could be performed in a decentralized (or distributed) way, meaning that sensors, together with the sink, cooperatively undertake the task of identifying the POI. However, unlike in classical decentralized detection problems, greater challenges exist in a WSN setting. There are stringent power constraints for each node, communication channels between nodes and the fusion center are severely bandwidth-constrained and are no longer lossless (e.g., fading, noise and, possibly, external sources of interference are present), and the observation at each sensor node is spatially varying. In the context of decentralized detection, cooperation allows the exchange of information among sensor nodes to continuously update their local decisions until consensus is reached across the nodes.

In SPE, the WSN aims at estimating a given physical phenomenon (e.g., the atmospheric pressure in a wide area, or the ground temperature variations in a small volcanic site), which can be modeled as a bi-dimensional random process (generally nonstationary). In this case, the main issue is to obtain the estimation of the entire behavior of the spatial process based on the samples taken by sensors that are typically placed in random positions [11]. The measurements will then subject to proper processing which might be performed either in a distributed manner by the nodes, or centrally at the supervisor. The estimation error is strictly related to nodes density as well as on the spatial variability of the process. Higher node density leads to a more accurate scalar field reconstruction at the expense of a larger network throughput and cost.

There exist also applications that belong to both categories. As an example, environmental monitoring applications could be ED- or SPE-based. To the first category belong, for example, the location of a fire in a forest, or the detection of a quake, etc. (see Fig. 8.5). Alternatively, the estimation of the temperature of a given area belongs to the second category. In general, these applications aim at monitoring indoor or outdoor environments, where the supervised area may be few hundreds of square meters or thousands of square kilometers, and the duration of the supervision may last for years. Natural disasters such as floods, forest fires, and earthquakes may be perceived earlier by installing networked embedded systems closer to places where these phenomena may occur. Such systems cannot rely on a fixed infrastructure and

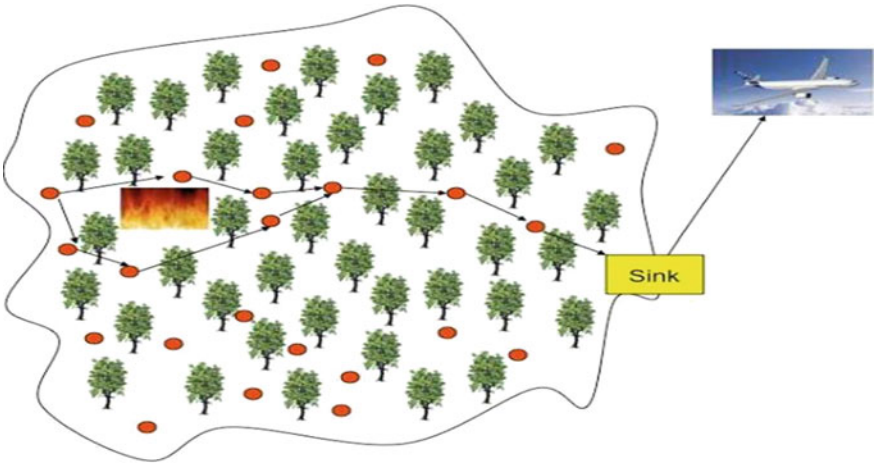


Fig. 8.5 Event detection application

have to be very robust, because of the inevitable impairments encountered in open environments.

The system should respond to environment changes as quick as possible. The environment to be observed will mostly be inaccessible by the human all the time. Hence, robustness plays an important role. Also, security and surveillance applications have some demanding and challenging requirements such as real-time monitoring and high security.

Another application that could belong to both the above-defined categories is devoted to the realization of energy-efficient buildings. In this application, in fact, sensor nodes could aim at estimating a process (SPE), but also events (ED). In this case, the WSN is distributed in buildings (residential or not) to manage efficiently the energy consumption of all the electric appliances. Consequently, nodes have to continuously monitor the energy consumed by all appliances connected to the electrical grid. Therefore, sensors have to estimate a process, that is the energy consumption which varies with time, but in some cases, they could be used to detect some events. As an example, sensors could detect the arrival of a person in a room to switch on some electrical appliances.

8.4.2 Examples of Application Requirements

Due to the wide variety of possible applications of WSNs, system requirements could change significantly. For instance, in the case of environmental monitoring applications, the following requirements are typically dominant: *energy efficiency*, nodes are battery powered or have a limited power supply; *low data rate*, typically

the amount of data to be sensed is limited; *one-way communication*, nodes act only as sensors and hence the data flow is from nodes to sink(s); *wireless backbone*, usually in environmental monitoring no wired connections are available to connect sink(s) to the fixed network.

Significantly different are the requirements of a typical industrial application where wireless nodes are used for cable replacement: *reliability*, communication must be robust to failure and interference; *security*, communication must be robust to intentional attacks; *interoperability*, standards are required; *high data rate*, the process to be monitored usually carries a large amount of data; *two-way communication*, in industrial applications nodes typically act also as actuators and hence the communication between sink(s) and nodes must be guaranteed; *wired backbone*, sinks can be connected directly to the fixed network using wired connections.

Even if requirements are strongly application dependent, one of the most important issues in the design of WSNs, especially in such scenarios where power supply availability is limited, is energy efficiency. High energy efficiency means long network lifetime and limited network deployment and maintenance costs. Energy efficiency can be achieved at different levels starting from the technology level (e.g., by adopting low consumption hardware components), physical layer, MAC, and routing protocols up to the application level. For example, at physical and MAC layers, nodes could operate with low duty cycle by spending most of their time in sleeping mode to save energy. This poses new problems such as that nodes may not wake up at the same time, due to the drifts of their local clocks, thus making the communication impossible.

The key requirements for transceivers in sensor networks are given in ZigBee.

- Low cost: Since a large number of nodes are to be used, the cost of each node must be kept small. For example, the cost of a node should be less than 1% of the cost of the product it is attached to.
- Small form factor: Transceivers' form factors (including power supply and antenna) must be small so that they can be easily placed in locations where the sensing actually takes place.
- Low energy consumption: A sensor usually has to operate for several years with no battery maintenance, requiring the energy consumption to be extremely low.

Some additional requirements are needed to make the wireless sensor network effective.

- Robustness: Reliability of data communication despite interference, small-scale fading, and shadowing is required so that high quality of service (e.g., with respect to delay and outage) can be guaranteed.
- Variable data rate: Although the required data rate for sensor networks is not as high as multimedia transmissions, low data rates may be adequate for simple applications while some other applications require moderate data rates.
- Heterogeneous networking: Most sensor networks are heterogeneous, i.e., there are nodes with different capabilities and requirements. Typically, the network has some full-function device (FFD) that collects data from different sensors,

processes them, and forwards them to a central monitoring station. An FFD has fewer restrictions with respect to processing complexity (as there are few FFDs, cost is not such an important factor) and energy consumption (since an FFD is usually connected to a permanent power supply). The sensor nodes themselves, on the other hand, are usually reduced function devices (RFDs) with extremely stringent limits on complexity and power consumption.

Apart from data communication, geolocation is another key aspect for many wireless sensor network applications. Normally, a number of nodes communicate their sensing (measurement) results to each other and/or a control center. In many cases, the control center or the receiving nodes need to know the exact location of the transmitter. For example, when a fire sensor detects the fire, the control center not only wants to know that there is a fire but also wants to know at which location. In a building automation system, a large number of sensors will be deployed with building equipment. Any detected abnormal condition along with its location will help the effort of diagnosis and maintenance significantly. Although some applications with geolocation needs may elect to manually enter the device's locations, many applications cannot afford either the time or cost associated with this practice. Location information is also important because monitoring and control systems often perform data analysis based on both spatial and temporal correlation from closely spaced sensors [12].

8.5 Characteristic Features of Wireless Sensor Networks

In ad hoc networks, wireless nodes self-organize into an infrastructure-less network with a dynamic topology. Wireless sensor networks share these traits, but also have several distinguishing features. The number of nodes in a typical sensor network is much higher than in a typical ad hoc network, and dense deployments are often desired to ensure coverage and connectivity; for these reasons, sensor network hardware must be cheap. Nodes typically have stringent energy limitations, which make them more failure-prone. They are generally assumed to be stationary, but their relatively frequent breakdowns and the volatile nature of the wireless channel nonetheless result in a variable network topology. Ideally, sensor network hardware should be power-efficient, small, inexpensive, and reliable in order to maximize network lifetime, add flexibility, facilitate data collection, and minimize the need for maintenance.

8.5.1 Lifetime

Lifetime is extremely critical for most applications, and its primary limiting factor is the energy consumption of the nodes, which need to be self-powering. Although it is

often assumed that the transmit power associated with packet transmission accounts for the lion's share of power consumption, sensing, signal processing, and even hardware operation in standby mode consume a consistent amount of power as well. Many researchers suggest that energy consumption could be reduced by considering the existing interdependencies between individual layers in the network protocol stack. Routing and channel access protocols, for instance, could greatly benefit from an information exchange with the physical layer.

At the physical layer, benefits can be obtained with lower radio duty cycles and dynamic modulation scaling (varying the constellation size to minimize energy expenditure). Medium Access Control (MAC) solutions have a direct impact on energy consumption, as some of the primary causes of energy waste are found at the MAC layer: collisions, control packet overhead, and idle listening. Energy-efficient routing should avoid the loss of a node due to battery depletion.

8.5.2 Flexibility

Sensor networks should be scalable, and they should be able to dynamically adapt to changes in node density and topology, like in the case of the self-healing minefields. In surveillance applications, most nodes may remain quiescent as long as nothing interesting happens. However, they must be able to respond to special events that the network intends to study with some degree of granularity. In a self-healing minefield, a number of sensing mines may sleep as long as none of their peers explodes, but need to quickly become operational in the case of an enemy attack. Response time is also very critical in control applications (sensor/actuator networks) in which the network is to provide a delay-guaranteed service.

8.5.3 Maintenance

The only desired form of maintenance in a sensor network is the complete or partial update of the program code in the sensor nodes over the wireless channel. The functioning of the network as a whole should not be endangered by unavoidable failures of single nodes, which may occur for a number of reasons, from battery depletion to unpredictable external events, and may either be independent or spatially correlated. Fault tolerance is particularly crucial as ongoing maintenance is rarely an option in sensor network applications. Self-configuring nodes are necessary to allow the deployment process to run smoothly without human interaction, which should in principle be limited to placing nodes into a given geographical area. Location awareness is important for self-configuration and has definite advantages in terms of routing and security. Time synchronization is advantageous in promoting cooperation among nodes such as data fusion, channel access, or security-related interaction.

8.6 Existing Technologies and Applications

Recently, most wireless sensor networks relied upon narrow-band transmission schemes such as direct sequence or frequency hopping along with multiple-access techniques such as carrier-sense multiple access (CSMA) carrier sense. For example, the narrow-band direct-sequence spread spectrum (DSSS) PHY layer that is currently used in conjunction with the ZigBee networking standard in the 2.4 GHz band 2 employs a 2 Mchip per second code shift keying modulation to provide 250 kbits/s. ZigBee can be used for wireless control and monitoring solutions without extensive infrastructure wiring. Wireless sensor networks using ZigBee can also be used to monitor logistics assets and track the objects. However, location estimation based on narrow-band DSSS can achieve accuracy on the order of several meters, which is only slightly more accurate than traditional RFID. The main initial markets of ZigBee are home, building, and industrial automation such as monitoring and control of lights and HVAC, security in commercial buildings and home, industrial monitoring and control, automatic meter reading, medical and health monitoring of patients, equipment, and facilities.

Other candidate technologies for WSNs are the various forms of IEEE 802.11 or WiFi. The IEEE ratified the initial IEEE 802.11 specification in 1997 as a standard for wireless Local Area Networks (WLANs). An early version of 802.11 (i.e., 802.11b) supports transmission up to 11 Mbits/s. Subsequent mainstream WLAN standards are 802.11a and 802.11g, which achieve 54 Mbits/s. Most recently, the 802.11n standard is under development to achieve more than 100 Mbits/s for high data rate applications and IEEE 802.11s is developed for realizing mesh networking. WiFi is designed for fast and easy networking of PCs, printers, and other devices in a local environment. It can provide much higher data rates than ZigBee with a longer communication distance per link. In addition, WiFi is a more mature technology and has been widely adopted in various applications. However, its complexity and energy consumption are much higher than that of ZigBee. For these reasons, WiFi technology has been applied only to perform some particular functions in wireless sensor networks. In many cases, it is used to collect sensor data for transmission over longer distance with fixed power supply. In some industrial and hospital wireless network systems, WiFi has also been used to monitor and locate facilities with an accuracy of several meters.

Compared to narrow-band DSSS and WiFi, UWB offers significant advantages with respect to robustness, energy consumption, and location accuracy. UWB spreads the transmit signal over a very large bandwidth (typically 500 MHz or more). By using a large spreading factor, higher robustness against interference and fading is achieved. The use of very short pulses in impulse radio transmission with careful signal and architecture design results in very simple transmitters and permits extremely low energy consumption. The average power consumption for UWB transceiver is about 30 mW which is similar to that of narrow-band ZigBee (20–40 mW) and much lower than 802.11g (500 mW–1 W).

Table 8.1 Comparison of wireless technologies

	2.4 GHz ZigBee	2.4 GHz WiFi	UWB
Data rate	Low, 250 kbps	High, 11 Mbps for 802.11b and 100 Mbps for 802.11n	Medium, 1 Mbit/s mandatory, and up to 27 Mbps for 802.15.4a
Transmission distance	Short, <30 m	Long, up to 100 m	Short, <30 m
Location accuracy	Low, several meters	Low, several meters	High, <50 cm
Power consumption	Low, 20–40 mW	High, 500 mW–1 W	Low, 30 mW
Multipath performance	Poor	Poor	Good
Interference resilience	Low	Medium	High with high complexity receivers, low with simplest receivers
Interference to other systems	High	High	Low
Complexity and cost	Low	High	Low–medium–high are possible

The precision of ranging measurements, which form the basis of geolocation, is proportional to the bandwidth that can be employed. Therefore, UWB also offers considerable advantages for geolocation with submeter accuracy. Better than 15 cm ranging accuracy and less than 50 cm location accuracy are achievable. Table 8.1 provides a comparison among the three abovementioned technologies [12].

8.7 Conclusions

Sensor networks offer countless challenges, but their versatility and their broad range of applications are eliciting more and more interest from the research community as well as from industry. Sensor networks have the potential of triggering the next revolution in information technology. The aim of this chapter is to discuss some of the most relevant issues of WSNs, from the application, design, and technology viewpoints. For designing a WSN, in fact, the most suitable technology is needed to define to be used and the communication protocols to be implemented (topology, signal processing strategies, etc.).

References

1. Culler D, Estrin D, Srivastava M (2004) Guest editors' introduction: overview of sensor networks. Computer 37(8):41–49
2. Nugroho DA, Prasetiadi A, Kim D-S (2014) Male-silkmoth-inspired routing algorithm for large-scale wireless mesh networks. J Commun Netw (JCN). IF: 0.747, ISSN: 1976-5541
3. Hoa TD, Kim D-S (2013) Data forwarding algorithm over lossy links in wireless sensor networks. IEICE Commun Expr 2(10):453–458. ISSN: 2187-0136
4. Verdone R (2008) Wireless sensor networks. In: Proceedings of the 5th European conference, Bologna, Italy
5. Basagni S, Conti M, Giordano S, Stojmenovic I (2004) Mobile ad hoc networking. Wiley, San Francisco, CA, USA
6. Akyildiz I, Su W, Sankarasubramaniam Y, Cayirci E (2002) A survey on sensor networks. IEEE Commun Mag 40(8):102–114
7. Tubaishat M, Madria S (2003) Sensor networks: an overview. IEEE Potentials 22(2):20–23
8. Lin C-Y, Tseng Y-C, Lai T (2006) Message-efficient in-network location management in a multi-sink wireless sensor network. In: International conference on sensor networks, ubiquitous, and trustworthy computing, vol 1, no 2, pp 8–14, June 2006
9. Tan DD, Kim D-S (2014) Dynamic traffic-aware routing algorithm for multi-sink wireless sensor networks. Wirel Netw 20(6):1239–1250. IF: 1.055, ISSN: 1572-8196
10. Hao Q, Brady D, Guenther BD, Burchett J, Shankar M, Feller S (2006) Human tracking with wireless distributed pyroelectric sensors. IEEE Sens J 6(6):1683–1696
11. Behroozi H, Alajaji F, Linder T (2008) On the optimal power-distortion region for asymmetric Gaussian sensor networks with fading. In: IEEE international symposium on information theory, July 2008, pp 1538–1542
12. Zhang J, Orlik PV, Sahinoglu Z, Molisch AF, Kinney P (2009) UWB systems for wireless sensor networks. IEEE

Chapter 9
Wireless Fieldbus for Industrial Networks

9.1 Introduction

The recent improvement of wireless communication networks has made possible for using such networks at the lowest levels of factory automation systems, which typically imposes severe requirements in terms of both real-time performance and dependability. However, industrial plants represent a hostile environment due to the presence of different noise sources. The presence of these noises can easily inflict damages on the wireless information signals and creates erroneous interruptions on the data transmitted. This chapter presents a survey and analysis of wireless fieldbus in industrial environments. For industrial environment, reliability, and tight timing requirements are difficult to satisfy by the specific properties of the wireless communications.

Nowadays, wired communication networks are utilized in industrial plant environment. The environment is known as the field level and the pertinent communication network is called Fieldbuses. In industrial network, the low level network connecting with sensors and actuators plays an important role. The general industrial environment is composed of a number of sensors and actuators. They send sensed data to base station by peer-to-peer. Fieldbuses create a real-time communication between the controller and sensors/actuators. At this level, there are typically transmission of limited amounts of data (tens of bytes or even less) between controllers and sensors/actuators deriving from two functions: cyclic data exchange and acyclic data handling. Fieldbus should be able to transmit a real-time periodic data for alarm and network maintenance [1]. By these limitations, most existing Fieldbus based on wired technology was introduced.

At this level, there is typically transmission of limited amounts of data (some tens of bytes or even less) between controllers and sensors/actuators deriving from two functions: cyclic data exchange and acyclic data handling. The cyclic data exchange accounts for the periodic polling executed by controllers on the field devices in order to transmit, for example, process values, set points, etc. In this case, the fundamental

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_9

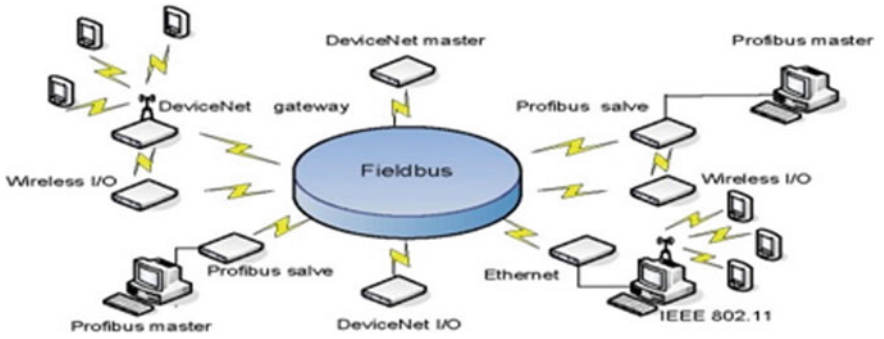


Fig. 9.1 Wireless Fieldbus

parameter is the jitter which is defined as the time variation with which the slaves are polled. The acyclic data handling function allows for the exchange of critical data such as, for example, those deriving from alarm situations. In this case, the fundamental parameter is the latency, defined as the time required for the actual transmission of the acyclic data to the destination. Both functions have to be performed with very tight timing constraints (at present, Fieldbuses may be requested to work with both periods and latencies in the order of milliseconds). Industrial environment represents a series of specific conditions such as high electromagnetic noises, recently, however, along with the impressive growth of wireless LAN, networks like these have become parts of the suitable communication systems. Thus, it is very likely that the use of wireless networks will be enhanced in factories communication systems in the near future. Owning to the recent advancements in wireless technology, the wireless network has been integrated to an existing Fieldbus system [2]. The Wireless Fieldbus has much strength on mobility, easy installation and maintenance. In general, the industrial environment should be error-prone and reliable. Therefore, it is important to design a wireless Fieldbus system that can support many nodes and guarantee real-time data transmission as shown in Fig. 9.1.

In Table 9.1, there are shown wired/wireless Fieldbus such as PROFIBUS, DeviceNet, R-Fieldbus, and vendor-specific protocol by Elpro technology [3]. There are a lot of suppliers that supply Wireless Fieldbus systems but each of them provides their own "type" of Wireless Fieldbus system, which means they have their own modem application, such as 802.11 WiFi Transparent modem, 869 MHz Fixed Frequency Transparent, Smart Radio Modem, and many more, etc. The manufacturers realized many different Fieldbus systems it disregarded rather than attracted the costumers. The companies then made their devices compatible with each other and they also made their specifications publicly available so different vendors can produce compatible devices.

There are some main differences between Wireless Fieldbus Systems:

- *Operation frequencies/bandwidth*: In the Wireless Fieldbus world, there are two different bandwidths, 900 MHz (902–928) and 2.4 GHz (2.4–2.4385). The

Table 9.1 Comparison between wired and wireless Fieldbus

Technology	Wired		Wireless	
Name	PROFIBUS	DeviceNet	R-Fieldbus	Elpro
Segment length (m)	1200 600 200	500 250 100	100	2000
Data rate (kbps)	93.75 182.5 500	125 250 500	2000	4.8
Nodes	32	64	30	95

900 MHz band is mostly used in the United States. Signal of a 900 MHz band can be divided into different specific frequencies. The 2.4 GHz band is the worldwide standard signals receive (RX) and transmit (TX) band designing from the WiFi system and can be used for communication between Ethernet and non-Ethernet device.

- *Standards*: IEEE 802.11—also known as WiFi or wireless Ethernet—is a family of standards for wireless local area networks. 802.11 technologies can be used to connect a laptop PC to a corporate or home network. In plant applications, it provides a cost-effective way to link small networks of wireless devices to a host system or plant Local Area Network (LAN). It can also be used for linking mobile workers' computers or Personal Digital Assistants (PDAs) to the LAN. The IEEE 802.15.4 radio standard provides a simple platform for low-cost, low-power, and high-reliability communication. This standard may provide the physical basis for process industry standards such as ISA-SP100, WirelessHART, or others. It uses two lower OSI layers including physical layer and data link layer. Its physical layer is a spread-spectrum radio that operates in the 2.4 GHz band at a rate of 250 kbps. Its medium-access control (MAC) layer supports the three common wireless topologies: star, mesh, and cluster-tree.
- *Network topology*: Wireless network topologies describe both the physical layout of devices and the routes that data follows for communication. There are several types of wireless topologies (shown in Fig. 9.2) and the most common wireless topologies are fully connected. The advantage of mesh and cluster-tree topology is that the transmitters can communicate to the receiver through each other. Thus, if there is an obstacle that breaks the signal between the transmitter and receiver, it can be delivered through another transmitter on the net.

Fig. 9.2 Different wireless topologies

9.2 Wireless Fieldbus Technology

9.2.1 Overview

The idea for the future is to make one standard for all of the Wireless Fieldbus systems. Different companies or consortia including companies, end users, and laboratories have developed the Fieldbus for many years. It has been impossible to find a real consensus at the international standardization level for several reasons. The first one is the number and the variety of the applications (process control, manufacturing systems, and automotive embedded systems, building automation systems, and so on). The second one is related to the different architecture principles to design and to implement the control system. A third reason, and maybe the most important, is the strategic aspect of the Fieldbus for the quality of service in a distributed system. For more details on this story, the interested readers may see [4–6].

Since the fast development of the Wireless Fieldbus systems, there are a lot of different standards. The most important standards which are used by the well-known Fieldbus manufacturers:

- The first proposal, produced by ABB, Emerson, Endress & Hauser, and Siemens, is based on the work of the so called "Heathrow Group". Their solution is going on WirelessHART protocol and determines some features of the ISA100.11a standard and to convergence with the Chinese WIA-PA standard. This would produce a single wireless standard, something the end users have been demanding for the past half-decade now.
- The second proposal, produced by Honeywell, Invensys, Nivis, Yokogawa, Fuji Electric, Hitachi America, and Yamatake, is to adopt the ISA100.11a as a standard.

- The third proposal is to rewrite ISA100.11a and make it compatible with WirelessHART, ISA100.11a, and WIA-PA.

Meanwhile, ISA made a new committee, the ISA100.12 committee. The idea is to include ZigBee and the new Chinese standard, as well as WirelessHART in a converged ISA100 standard [7].

9.2.2 Wireless Fieldbus Systems Proposals

As described in [8], the approaches differ in the way devices are interconnected. In the simplest case, all devices belong to a single wireless cell and no connection is made to wired segments. Meanwhile, a wired segment and a wireless cell, in the repeater approach, are linked together by a repeater that converts signals bit by bit, or character by character, from one medium to the other [9]. In the bridge approach, a different medium-access control is used on the wire segment abandon the wireless cell [10] and in the gateway approach, a special device with the gateway, allows to "see" the set of wireless devices as if they were connected to the wired segment [11].

9.2.2.1 R-Fieldbus

On January 2000, the IST project "High Performance Wireless Fieldbus in Industrial-related Multimedia Environment (R-FIELDBUS)" has been started. The consortium set oneself the target to develop a high-performance Wireless Fieldbus architecture, providing data rates of up to 2 Mbit/s and response times similar to wired Fieldbus solutions. To achieve this target, new high-performance radio technologies and existing industrial communication protocols must be integrated to provide a flexible wireless Fieldbus architecture. This architecture must be able to handle with the real-time necessities of the distributed control data, to support a user-defined Quality of Service (QoS) concerning industrial multimedia services, to support mobility of devices and to support interoperability with existent communication infrastructures. The main actions must be undertaken for the design, development, implementation, and demonstration of the proposed R-FIELDBUS system.

The R-Fieldbus architecture is one of the examples on the integration of emerging wireless technologies for broadband systems and networks with existent industrial communication protocols such as those specified in the European Standard EN50170 [12]. The R-Fieldbus system must provide full transparent access to any information needed on site, such as data concerning real-time control and status information, or transparent to specification drawings and other industrial-type multimedia information (real-time voice and low-resolution digital video sequences). The R-Fieldbus support mobile industrial devices and its interoperability with existing devices supported by wired industrial networks [13]. The main subsystems of the R-Fieldbus system are:

- A high-speed, high-performance and reliable radio physical layer for important industrial requirements such as real-time and reliability.
- A data link layer coexisting with the necessary protocol extensions for the additional services (real-time distributed control traffic, industrial-related multimedia traffic, etc.).
- A device which allows the interconnection of the R-Fieldbus arrangement to Wireless Fieldbus networks in the industrial environments.
- A QoS mechanism to guarantee the required quality of service to the multimedia communication services.
- A set of application-level sustenance services covering issues such as the interchange of hard real time control data, interchange of industrial multimedia traffic data and mobile IP, giving transparent access to manufacturing and management information systems.
- Advanced network management services to support the real-time requirements and the required robustness of the Wireless Fieldbus, with fault tolerance, security, and safety mechanisms. The R-Fieldbus system can be tested and evaluated within two different field trials based on pilot applications. These trials can demonstrate both the R-Fieldbus technical feasibility in real industrial environments and its benefits for the end user applications.

9.2.2.2 Industrial WLAN

In recent studies and research, the IEEE 802.11 standard for wireless local area networks, has the best capability for the industrial applications. IEEE 802.11 standard is suitable only for the lower layers of the communication stack; however, it is necessary to complete the stack with appropriate protocols in industrial communication. In [14], the researchers had explored the use of IEEE 802.11 in industrial communication by analyzing the possibility of implementing protocols, based on Master–Slave architecture of wired Fieldbus on IEEE 802.11 Physical and Data Link Layers. In [15], the authors used some measurement sets of IEEE 802.11 wireless link in an industrial environment to conclude that the behavior of the wireless link is characterized by time-varying behavior of wireless channel, packet losses, and bit errors. In [16], all abovementioned architectures have been used a simple polling scheme for the exchange of cyclic data and have considered three different techniques (Late, Current, and Immediate) for handling acyclic request in AWGN channel. However, IEEE 802.11 node has high cost and reliability limitation. In general, many nodes should be allocated in industrial environment. Therefore, the Wireless Fieldbus based on IEEE 802.11 has problems because of high installing and maintenance cost.

9.2.2.3 IEEE 802.15.4 (LR-WPAN)

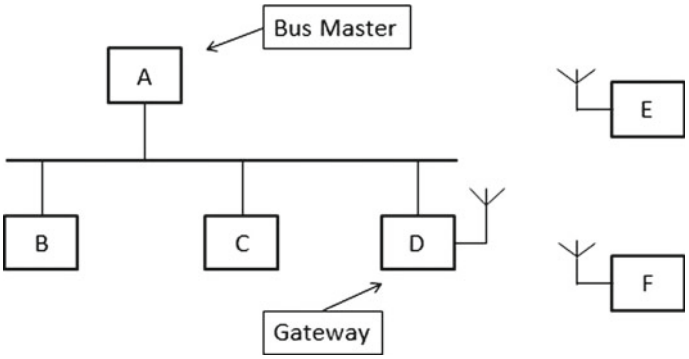


Fig. 9.3 FIP network

The IEEE 802.15.4 is called a low-rate wireless personal area network (LR-WPAN). The LR-WPAN was proposed for data gathering and control through sensors. The coordinator node of LR-WPAN can manage many nodes by mesh network [17]. Each node in LR-WPAN has low price and a coordinator node can support many I/O nodes. However, it can't handle mixed real-time traffic efficiently. Therefore, efficient transmission of mixed real-time traffic has been required.

9.2.2.4 Factory Information Protocol (FIP)

The Laboratory of Industrial Computing is active in the study of the Fieldbus protocols including the FIP (Factory Information Protocol) protocol [18]. In this context, a need has been identified to provide wireless access to distant sensors and actuators through a FIP Fieldbus, using inexpensive wireless modems which are commercially available. The ideal situation would be that the distant sensors and actuators be integrated into the FIP network in a totally transparent manner, so that a station of the FIP network be given access to the distant stations as for any other station in the network, whether local or wireless connected (as shown in Fig. 9.3). Unfortunately, the bandwidth and error rates of available wireless connections would significantly degrade the Fieldbus performance. FIP, in particular, requires that polling sequences be answered within 70-bit times from their reception. Another problem is that FIP operates in broadcast mode, where each station listens to all others and copies the variables that it needs as they are transmitted from their producers. In an industrial environment it may be extremely difficult to provide any-to-any connectivity among wireless stations, it is much easier to arrange for one-to-any connectivity around a base station. Figure 9.3 shows an example of a FIP network on which are connected a bus master A, fixed stations with sensors/actuators, B, C, a gateway D, and wireless stations E and F. The wireless stations exchange variables transparently through the gateway.

Morel et al. have studied several solutions to integrate wireless nodes to a FIP Fieldbus [19]. The first solution is a simple repeater that converts from cable to radio and vice versa. The various ways to handle FIP constraints are discussed. The second solution [20] called "word repeater" approach in which they use all frames coming from the wired segment are repeated byte by byte with additional forward error correcting code in order to increase to reduce the bit error rate on the wireless side. They used Golay code which seems to be the best solution to overcome the limitations due to the wireless transceiver switching time.

9.2.2.5 Other Wireless Fieldbus Systems

Fieldbus international standard has defined a radio physical layer that can be used in place of wireline transceivers. In addition, there are a few individual products available that use some sort of wireless link. However, none of these products is a standard. Almost research on Fieldbus, at its beginning, was focused on the definition of the services and of new protocols to provide the right quality of service for given kinds of application, and for the given distributions of the control systems. For some years, this research has been oriented toward the quality of service and toward the design of system architectures. And currently, there are two main areas of interest. The first one is related to the development of applications and the second one is dedicated to the definition of new services or new protocols. Several associated topics include the development of the use of new transmission media (as wireless) but also the definition of new traffic management policies.

The paper presented by Juanole [21] focus on the combined modeling of the communication stack and of the application processes in order to evaluate the quality of service viewed by the end user, in terms of application. Other papers presented by Simonot-Lion and Elloy [22, 23] are dedicated to an Architecture Description Language for the development of Fieldbus based distributed and embedded systems. This language may be considered as an interesting proposal for the application specifications. The papers presented by Neumann [24, 25] focus on the device description models and languages, their interest for the interoperability of Fieldbus based systems and their necessity for the integration in management systems. The papers presented by Jean-Dominique [26] and Decotignie [27] are devoted to a state of the art on Wireless Fieldbus. Salvatore Cavalieri and Salvatore Monforte presented a profile modeling for verification of constraints by performance evaluation [28].

9.3 Issues in Wireless Fieldbus Networks

Wireless Local Area Network (WLAN) and the Wireless PAN standard (Bluetooth and ZigBee) utilize the unlicensed 2.4 GHz ISM band. Due to their dependence on the same band, the potential for interference exists. In [29], the experiments and measurements were evaluated to qualify the interference effect of Bluetooth devices

on the throughput performance of IEEE 802.11g and 802.11b. The results show how 802.11g is immune to interferences than 802.11b when the signal strength of the WLAN is strong. In [30], the Packet Error Rate (PER) of IEEE802.15.4 low-rate WPAN under the interference of IEEE802.11b WLAN was analyzed. The Bit Error Rate (BER) of IEEE 802.15.4 is obtained from the offset quadrature phase shift keying modulation. The bandwidth of IEEE802.11b is larger than that of IEEE 802.15.4, the in-band interference power of IEEE 802.11b is considered as the additive white Gaussian noise for the IEEE 802.15.4. Due to collision, time is calculated under the assumption that both packet transmissions are independent. If the distance between IEEE 802.15.4 and WLAN is longer than 8 m, the interference of WLAN does not affect the performance of the IEEE 802.15.4. If the frequency offset is higher than 7 MHz, the interference influence of the WLAN is negligible to the performance of the IEEE 802.15.4. Finally, three additional channels of the IEEE 802.15.4 on 2260, 2350, and 2470 MHz can be used for the coexistence channels under the interference of the between IEEE 802.15.4.

9.3.1 Consistency Problems of Fieldbus Technology

When a wireless control network arrangement uses the producer distributor–consumer communication model, which wireless Fieldbus ensures, communication is centered on an acknowledged broadcast of data identifiers to which the station possessing the identified data item broadcasts its actual value. If wireless channel is same for every transmitter–receiver pair a packet is corrupted independently with a firm possibility. When the producer obtains the identifier and broadcasts the data value, triumph spatial reliability requires that consumers obtain the data packet, which chances with probability. To attain comparative temporal reliability, all the sensors must sample the process within the same prespecified time window.

9.3.2 Problems for Token-Passing Protocols

Wireless Fieldbus arrangements similar to the PROFIBUS rely on distributed token passing in order to circulate the right to initiate transmissions between numbers of controllers. The master stations area ranged in a logical ring on top of a transmission medium that repetitive losses of token packets are unadorned problem for the constancy of the logical ring. Token-passing protocols thus serve as an example for the fact that it is often not only the raw presence of bit errors that is important, but also the characteristics of the faults as well. Improvement is that the communicating right can be granted to each station. Drawback is that when load is small, the token may go around, in effectiveness occurs. Since there are losses of the token or repetition of tokens, a system must become convoluted.

9.3.3 Problems in CSMA Based Protocol

Fieldbus systems on CAN use CSMA based protocols where collisions are possible. CSMA-based protocols work in a scattered way, where a station requiring to transmit first needs to sense the transmission medium. If the medium is determined to be idle, the station begins to transmit. The many Carrier Sense Multiple Access (CSMA) variants that exist differ in what happens when sensing the medium bus. The CAN protocol is grounded on a deterministic mechanism to determinate this argument. However, this mechanism is difficult to use for wireless media. It trusts on a stations ability to transmit and receive concurrently on the same channel, which is impossible with half duplex wireless transceivers.

9.4 Conclusions

In this chapter, an overview of the problems and issues about the application of wireless technologies on Wireless Fieldbus technology has been presented. The modeling of Fieldbus with the objectives of proof and of performance evaluation is of a major interest for the quality of the application design result. All the papers of this session are related to this main problem. Different proposals in the industrial domain and solutions are described. At this point, we lack a solution that covers all requirements especially because existing wired solutions cannot be readily expanded. It is hence difficult to offer the required reliability and temporal guarantees unless the environment can be protected. This calls for solutions using protected bands. With the explosion of projects in the area of wireless sensor networks, we may see good solutions in the near future.

References

1. Kim D-S, Lee YS, Kwon WH, Park HS (2003) Maximum allowable delay bounds of networked control systems. Control Eng Practice 11(11):1301–1313
2. Willig A, Matheus K, Wolisz A (2005) Wireless technology in industrial networks. Proc IEEE 93(6):1130–1151
3. Matkurbanov P, Lee S, Kim D-S (2006) A survey and analysis of wireless Fieldbus for industrial environments. In: SICE-ICASE, international joint conference, IEEE, 2006, pp 5555–5561
4. Thomesse J-P (2002) Fieldbuses and quality of service. In: 5th Portuguese conference on automatic control-controlo, pp 10–14
5. Fantoni D (1999) A never-ending story: the Fieldbus standardization. Automazione e Strumentazione 47(11):69–75
6. Thomesse JP (1999) Fieldbuses and interoperability. Control Eng Pract 7(1):81–94
7. https://www.bluetooth.org. Bluetooth standard
8. Decotignie J-D (2002) Wireless Fieldbusses–a survey of issues and solutions. In: Proceedings 15th IFAC world congress on automatic control (IFAC), 2002

9. Morel P, Croisier A, Decotignie J-D (1996) Requirements for wireless extensions of a FIP Fieldbus. In: Emerging technologies and factory automation, EFTA'96. Proceedings, vol 1, pp 116–122
10. Cavalieri S, Panno D (1998) A novel solution to interconnect Fieldbus systems using IEEE Wireless Lan technology. Comput Standards Interfaces 20(1):9–23
11. Morel P, Croisier A (1995) A wireless gateway for Fieldbus. In: Sixth IEEE international symposium on, vol 1. IEEE, 1995, pp 105–109
12. Thomesse JP (2000) Radio communication in automation systems: the R-Fieldbus approach. In: IEEE international workshop on factory communication systems, 2000, pp 319–326
13. Rauchhaupt L (2002) System and device architecture of a radio based Fieldbus—the R-Fieldbus system. In: IEEE international workshop on factory communication systems, 2002, pp 185–192
14. De Pellegrini F, Miorandi D, Vitturi S, Zanella A (2006) On the use of wireless networks at low level of factory automation systems. IEEE Trans Ind Inform 2(2):129–143
15. Miorandi D, Vitturi S (2004) Analysis of master-slave protocols for real-time industrial communications over IEEE 802. 11 WLANS. In: 2nd IEEE international conference on industrial informatics, INDIN'04, 2004, pp 143–148
16. Willig A, Kubisch M, Hoene C, Wolisz A (2002) Measurements of a wireless link in an industrial environment using an IEEE 802.11—compliant physical layer. IEEE Trans Ind Electron 49(6):1265–1282
17. Choi D-H, Lee JI, Kim D-S, Park WC (2006) Design and implementation of wireless Fieldbus for networked control systems. In: SICE-ICASE, IEEE international joint conference, 2006, pp 1036–1040
18. Neumann P (2007) Communication in industrial automation what is going on? Control Eng Prac 15(11):1332–1347
19. Haehniche J, Rauchhaupt L (2000) Radio communication in automation systems: the R-Fieldbus approach. In: IEEE international workshop on factory communication systems, 2000, pp 319–326
20. Winance M (2007) Being normally different changes to normalization processes: from alignment to work on the norm. Disabil Soc 22(6):625–638
21. Juanole G (2002) Quality of service of communication networks and distributed automation: Models and performances. In: Proceedings of 15th triennial IFAC world congress, 2002
22. Elloy J-P, Simonot-Lion F et al. (2002) An architecture description language for in-vehicle embedded system development. In: 15th IFAC world congress, Barcelona, Spain, 2002
23. Simonot-Lion F (1999) Une contribution a la modelisation et´ a la validation d'architectures temps reel, HDR thesis, INPL, Nancy, France, 1999
24. Neumann P, Diedrich C, Simon R (2002) Engineering of field devices using device descriptions. In: Proceedings IFAC world congress, 2002
25. Neumann P (1999) Locally distributed automation–but with which Fieldbus system? Assembly Autom 19(4):308–312
26. Decotignie J-D (1999) Some future directions in Fieldbus research and development, In: Fieldbus technology, Springer, Berlin, pp 308–312
27. Cucej Z, Gleich D, Kaiser M, Planinsic P (2004) Industrial networks. In: 46th international symposium electronics in Marine, 2004, pp 59–66
28. Cavalieri S, Monforte S, Tovar E, Vasques F (2002) Multi-master profibus-DP modelling and worst-case analysis based evaluation. In: Proceedings of the 15th IFAC world congress on automatic control, Barcelona, Spain, 2002
29. Partner R, THOB T, NTU D. D B3. 5-specification of residential gateway enhancements
30. Feuerstein MJ, Blackard KL, Rappaport TS, Xia H (1994) Path loss, delay spread, and outage models as functions of antenna height for microcellular system design. IEEE Trans Veh Technol 43(3):487–498

Chapter 10
Wireless Sensor Networks for Industrial Applications

10.1 Introduction

In the past few years, wireless sensor networks (WSNs) technology has been successfully applied in military, environment, health, home, and other commercial areas [1]. Compared to traditional wired device condition monitoring and diagnosis systems, using industrial wireless sensor networks (IWSNs) has inherent advantages, such as relatively low cost, and convenience of installation and relocation. By replacing periodic manual checkups with continuous wireless monitoring, it is claimed that industries could save up to 18% of the energy consumed by motor systems. So, industrial monitoring systems based on IWSNs have many potential advantages, although they have been relatively unexplored until recently.

This chapter focuses on the use of WSN in industrial applications also referred to as Industrial Wireless Sensor Networks (IWSN) and specifically considers industrial applications for control systems, which are different from the conventional control systems [2]. The industrial segment is an ever growing sector and huge amounts of capital are invested on research activities to support the advancements in WSN technology.

A general centralized IWSN scenario is depicted in Fig. 10.1 with nodes, sink/network manager, management console, and process controllers. The nodes collect data and communicate it to the sink/network manager which in turn communicates this data to the process controller. The nodes are managed by the network manager and the network manager can be controlled via a management console. The black arrows show a path through which a sensor node at the far end communicates to the sink via other nodes. In the control automation segment of the industry, the use of WSN as a part of the control loop has given rise to new possibilities. In these types of networks, the process controllers (actuators) are a part of the sensor network as shown in Fig. 10.2.

The nodes communicate data directly to the actuators (dashed arrows) and the actuators may also have some communication among themselves (solid arrows).

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_10

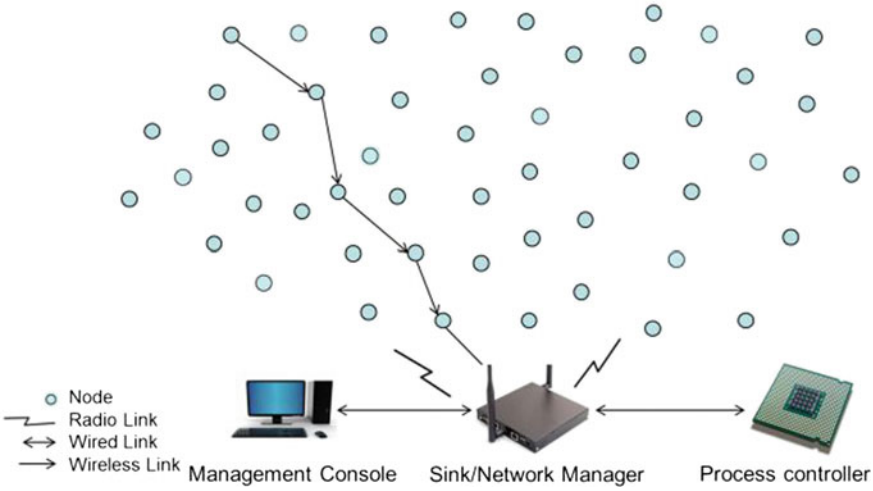


Fig. 10.1 General IWSN

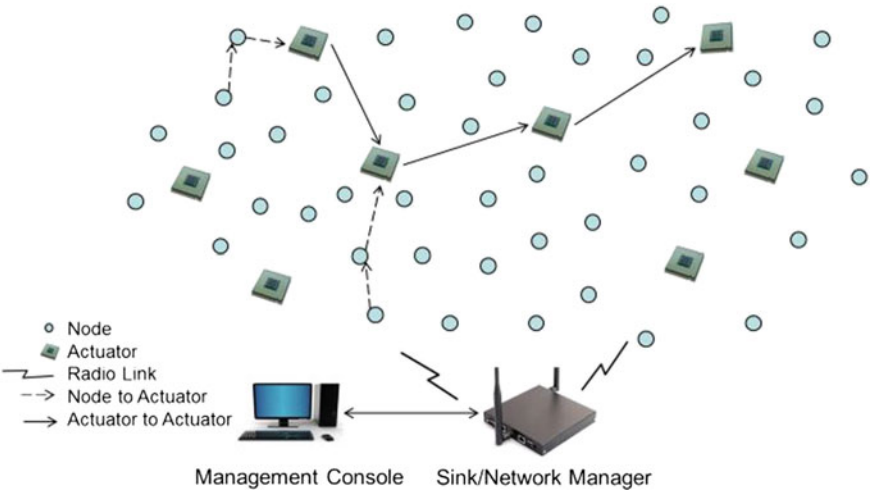


Fig. 10.2 Wireless sensor/actuators networks

These networks are referred to as wireless sensor and actuator (actor) networks (WSAN). The actuators are used to operate units, e.g., a valve and this is done based on the data sent by the sensors, e.g., temperature and pressure.

There are various wireless standards proposed to be used in IWSN, most importantly, ZigBee [3], ISA100.11a [4, 5], and WirelessHART [6]. These standards are used in various IWSN applications and are designed like frameworks that can be customized to the needs of a particular application setting, since they are not strictly defined at each layer of the communication protocol stack.

10.2 Industrial Wireless Sensor Networks

According to the International Society of Automation, the industrial systems can be classified into six classes [7] based on criticality of data and operational requirements. These classes range from critical control systems to monitoring systems, and their operational requirements and criticality vary accordingly.

10.2.1 Safety Systems

Systems where immediate (in the order of ms or s) action on events is required in the order of seconds, belong to this class, e.g., fire alarm systems. The WSN nodes are deployed uniformly throughout the area of concern to cover the entire area. The nodes are usually stationary.

10.2.2 Closed-Loop Regulatory Systems

Control system where feedbacks are used to regulate the system. WSN nodes are deployed in the area of concern in a desired topology. Periodically and based on events, measurements are sent to the controller. Periodic measurements are critical for the smooth operation of the system. These systems may have timing requirements that are stricter than safety systems. Based on these measurements, the controller makes a decision and sends it to the actuators which act on this data. Due to its strict requirements, a new protocol suite is proposed for this class of systems [8]. A simple control loop with wireless sensors and an actuator is shown in Fig. 10.3.

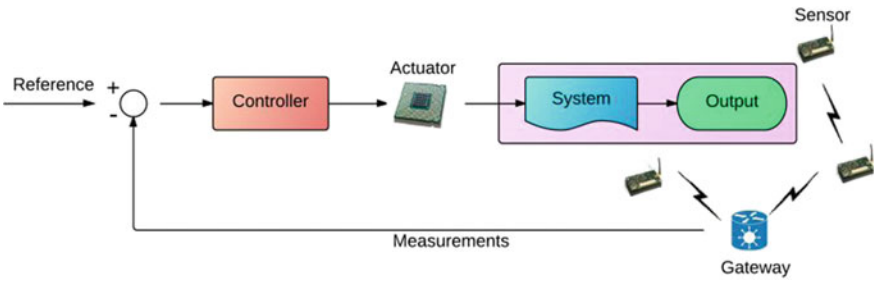


Fig. 10.3 Wireless closed-loop control with sensors and actuators

10.2.3 Closed-Loop Supervisory Systems

Similar to regulatory systems with the difference that feedbacks/measurements are not expected periodically but can be based on certain events. The feedbacks are noncritical, e.g., a supervisory system that collects statistical data and reacts only when certain trends are observed, which can be related to an event.

10.2.4 Open Loop Control Systems

Control systems operated by a human operator, where a WSN is responsible for data collection and relaying the collected data to the central database. The operator analyzes this data and undertakes any measures if required.

10.2.5 Alerting Systems

Systems are with regular/event-based alerting. An example is a WSN for continuous monitoring of temperature in a furnace and alerting at different stages to indicate part of the work done.

10.2.6 Information Gathering Systems

System used for data collection and data forwarding to a server. An example could be WSN nodes deployed in a field to gather data about the area of interest, such as temperature and moisture, for a specific duration of time. This data gathered over a long period can then be used to decide on long-term plans for managing temperature and moisture.

10.3 Industrial Standards

IWSN are required to have some basic qualities: low power, high reliability, and easy deployment, administration, and maintenance. These basic requirements drive the design goals for these devices. Various working groups like the Wireless Networking Alliance (WINA), the ZigBee, Alliance, the HART Communication Foundation (HCF), the International Society of Automation, and the Chinese Industrial Wireless Alliance have established standards for IWSN. The resulting standards are wirelessHART, ZigBee,, and ISA100.11a which is all based on the IEEE 802.15.4 standard. In this chapter, we discuss ZigBee,, wirelessHART, and ISA100.11a project.

10.3.1 ZigBee

ZigBee is one of the technologies that support the communication media. ZigBee is specification for a suite of high-level communication protocol. ZigBee technology is a low data rate, low power consumption, low cost, and wireless network protocol targeted toward automation and remote control application. It focused on Radio Frequency (RF) application which needed low power consumption, long-lasting battery, and secure network.

ZigBee classified standard IEEE 802.15 family with Bluetooth (802.15.1) and UWB (802.15.3) with standard code IEEE 802.4. The maximum communication speed about 250 kbps, range of communication 10–70 m, and using three band frequencies: 915 MHz (America), 868 MHz (Europe), 2.4 GHz (Japan). ZigBee, and Bluetooth are in Wireless Personal Area Network (WPAN) family. The differences between ZigBee and Bluetooth are in data rate, range, and Quality of Service (QoS). Based on the working principle, ZigBee is making full use of advantages from physical radio which very useful from standard IEEE 802.15.4. It is adding the logic network, security system, and application software.

ZigBee devices are of three types:

ZigBee Coordinator (*ZC*): The most capable device, the Coordinator forms the root of the network tree and might bridge to other networks. There is exactly one ZigBee Coordinator in each network since it is the device that started the network originally (the ZigBee Light Link specification also allows operation without a ZigBee Coordinator, making it more usable for over-the-shelf home products). It stores information about the network, including acting as the Trust Center and repository for security keys.
ZigBee Router (*ZR*): As well as running an application function, a Router can act as an intermediate router, passing on data from other devices.
ZigBee End Device (*ZED*): Contains just enough functionality to talk to the parent node (either the Coordinator or a Router); it cannot relay data from other devices. This relationship allows the node to be asleep a significant amount of the time thereby giving long battery life. A ZED requires the least amount of memory, and therefore can be less expensive to manufacture than a ZR or ZC.

The current ZigBee protocols support beacon and non-beacon-enabled networks. In non-beacon-enabled networks, an unslotted CSMA/CA channel access mechanism is used. In this type of network, ZigBee Routers typically have their receivers continuously active, requiring a more robust power supply. However, this allows for heterogeneous networks in which some devices receive continuously, while others only transmit when an external stimulus is detected. The typical example of a heterogeneous network is a wireless light switch: The ZigBee node at the lamp may receive constantly, since it is connected to the mains supply, while a battery-powered light switch would remain asleep until the switch is thrown.

The switch then wakes up, sends a command to the lamp, receives an acknowledgment, and returns to sleep. In such a network the lamp node will be at least a

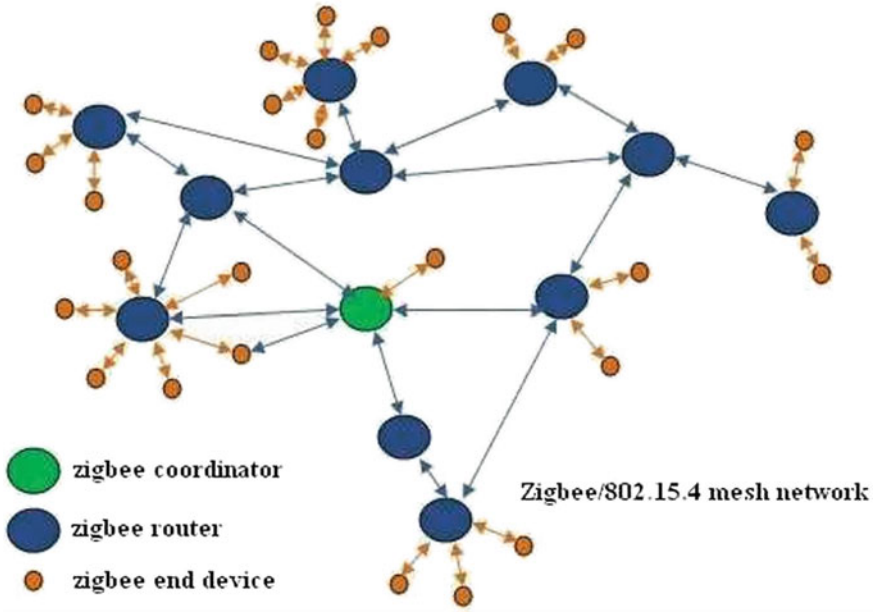


Fig. 10.4 ZigBee network overview

ZigBee Router, if not the ZigBee Coordinator; the switch node is typically a ZigBee End Device. In beacon-enabled networks, the special network nodes called ZigBee Routers transmit periodic beacons to confirm their presence to other network nodes. Nodes may sleep between beacons, thus lowering their duty cycle and extending their battery life. Beacon intervals depend on data rate; they may range from 15.36 ms to 251.65824 s at 250 kbit/s, from 24 ms to 393.216 s at 40 kbit/s, and from 48 ms to 786.432 s at 20 kbit/s. However, low duty cycle operation with long beacon intervals requires precise timing, which can conflict with the need for low product cost.

As mentioned in the network diagram, ZigBee network is comprised of coordinator (ZC), ZigBee router (ZR) and end ZigBee devices (ZE) as shown in Fig. 10.4.

10.3.2 WirelessHART

This standard is based on the IEEE 802.15.4 physical layer, with an operation frequency of 2.4 GHz and uses 15 different channels. It uses the Time Synchronized Mesh Protocol (TSMP) which was developed by Dust Networks for medium access control and network layer functions. TSMP uses TDMA for channel access and allows for channel hopping and channel blacklisting at the network layer. Channel hopping is a technique in which data transfer happens at different frequencies at different periods of time. The WirelessHART standard supports up to 15 channels

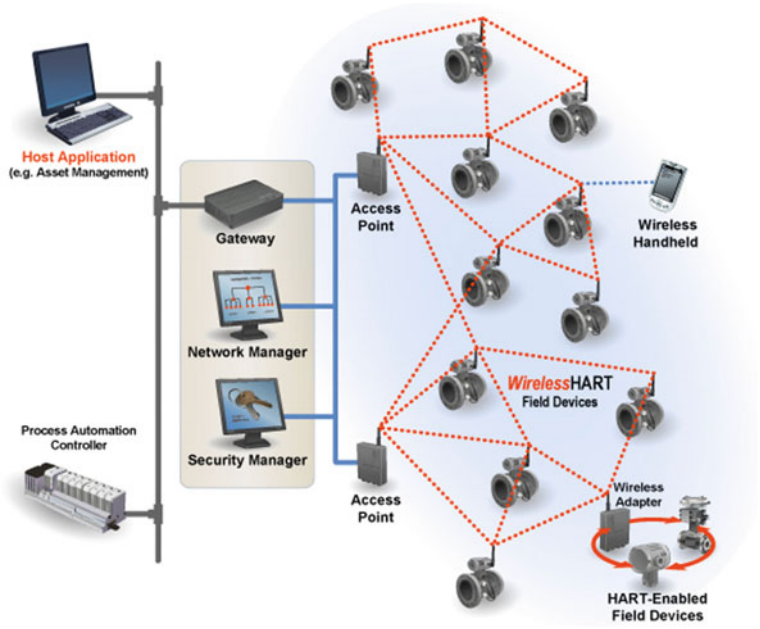


Fig. 10.5 WirelessHART network overview

which are used in turns. Channel blacklisting is a process of blacklisting channels which exhibit large interference with the signals. This use of TDMA with channel hopping and channel blacklisting has decreased the effect of interference and noise. WirelessHART supports redundant routing in order to enhance reliability. WirelessHART is thus considered to be robust, energy efficient and reliable, but since this is still an emerging standard, there is a lot of scope for improvement. WirelessHART was designed, developed, and standardized with industrial systems in mind and supports legacy systems built on wired HART. The network topologies supported by the network manager in WirelessHART are Star and Mesh.

WirelessHART is a wireless mesh network communications protocol for process automation applications. It adds wireless capabilities to the HART Protocol while maintaining compatibility with existing HART devices, commands, and tools as Fig. 10.5.

Each WirelessHART network includes three main elements:

- Wireless field devices connected to process or plant equipment. This device could be a device with WirelessHART built in or an existing installed HART-enabled device with a WirelessHART adapter attached to it.
- Gateways enable communication between these devices and host applications connected to a high-speed backbone or other existing plant communications network.
- A Network Manager is responsible for configuring the network, scheduling communications between devices, managing message routes, and monitoring network

health. The Network Manager can be integrated into the gateway, host application, or process automation controller.

The network uses IEEE 802.15.4 compatible radios operating in the 2.4 GHz Industrial, Scientific, and Medical radio band. The radios employ direct-sequence spread spectrum technology and channel hopping for communication security and reliability, as well as TDMA synchronized, latency-controlled communications between devices on the network. This technology has been proven in field trials and real plant installations across a broad range of process control industries. Network manager determines the redundant routes based on latency, efficiency, and reliability. To ensure the redundant routes remain open and unobstructed, messages continuously alternate between the redundant paths. Consequently, like the internet, if a message is unable to reach its destination by one path, it is automatically rerouted to follow a known-good, redundant path with no loss of data.

10.3.3 ISA100.11a

ISA100 working group developed this standard in order to provide robust and secure communication for applications in process automation [9]. Similar to wirelessHART, the physical layer is based on IEEE 802.15.4. ISA100.11a also uses channel hopping and channel blacklisting to reduce interference effects. ISA100.11a applies different methods for channel hopping like slow hopping, fast hopping, and mixed hopping. At the data link layer, it combines TDMA with CSMA in order to capitalize on the advantages in both solutions.

In general, a completed ISA100.11a-compliant system consists of three parts: the Data Link (DL) subnet, the Backbone Network (BN), and the Manager Network (MN), as depicted in Fig. 10.6. The DL subnet is comprised of I/Os and routing devices as well as portable devices. The backbone network includes backbone routers, and gateways (GWs). Finally, the MN consists of a system manager and a security manager. The ISA100.11a-compliant system is a centralized network governed by the MN. Thanks to information collected from the entire network, the MN provides a routing algorithm then load routing table to each device in the DL subnet. Data from field devices may transform directly to GW (one hop) or by the assistance of routers (multi-hop), as decided by MN.

The physical layer of ISA100.11a devices uses IEEE 802.15.4-2006radio transmission hardware. In fact, ISA100.11a uses only 16 channels (from 11 to 26), which operate at a maximum speed of 250 kbps with channel hopping. During operation, the timing axis of each device is broken down into slots typically varying from 10 to 12 ms. This timeslot duration in a DL subnet is set to a specific value by the MN when a device joins the network.

To be compatible with WirelessHART, 10 ms is a typical value for timeslot duration; however, duo casts could involve increased timeslot duration of approximately 1–2 ms. With a total of 16 channels, ISA100.11a topology supports three general

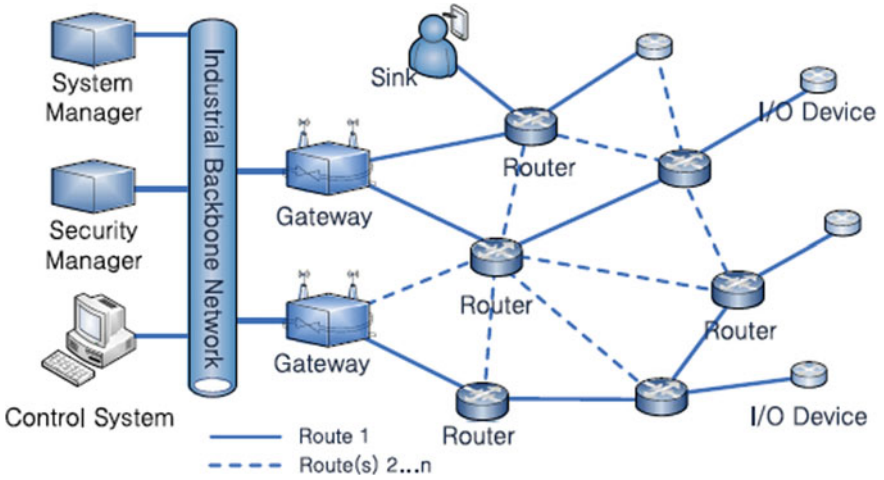


Fig. 10.6 ISA100.11a mesh network with backbone

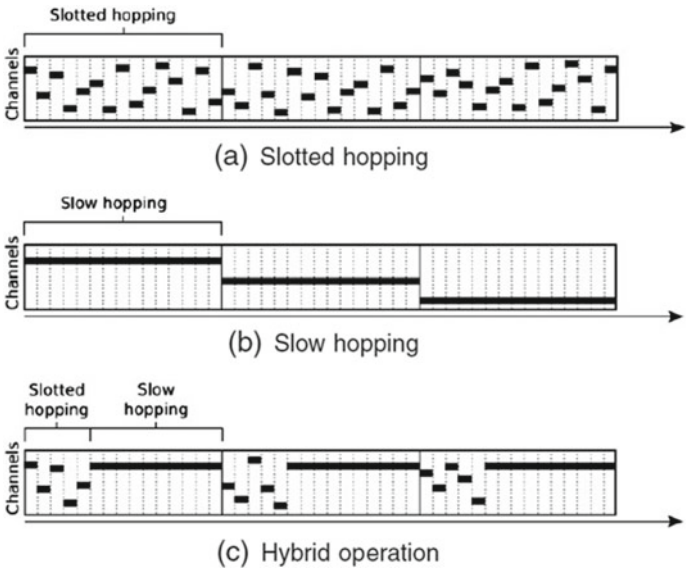


Fig. 10.7 Three alternative operations of one ISA100.11a subnet: **a** slotted hopping, **b** slow hopping, and **c** hybrid operation

alternative operations within one subnet: slotted channel hopping, slow channel hopping, and a hybrid method combining the first two methods, as depicted in Fig. 10.7. With the slotted hopping method, each timeslot uses a different radio channel to accommodate single transaction (with optional acknowledgment).

At the network layer, the compatibility with IPv6 gives opportunities for users to connect to the Internet, thus providing diverse possibilities. The ISA standard supports integration with legacy protocols like wired HART. In addition, the ISA also provides interface for facilitating co-existence with WirelessHART.

Based on the summary of state-of-the-art wireless standards we discuss some recent advances and current market share. Among all these standards, wirelessHART and ISA100.11a is the two major and dominating standards already in the market. In spite of the competition, the Hart Communication Foundation (HCF) and International Society of Automation (ISA) have agreed to collaborate together to produce one single standard derived from wirelessHART and ISA100.11a. A subcommittee named ISA100.12 has been created to investigate the possibilities of convergence. The convergence could result in a global standard with positives of both these standards and improved IWSN solutions.

10.4 Wireless Sensor Networks for Industrial Applications

WSNs can be used advantageously for rare event detection or periodic data collection for industrial applications. In rare event detection, sensors are used to detect and classify rare, random, and ephemeral events, such as alarm and fault detection notifications due to important changes in machine, process, plant security, operator actions, or instruments that are used intermittently. On the other hand, periodic data collection is required for operations such as tracking of the material flows, health monitoring of equipment/process. Such monitoring and control applications reduce the labor cost, human errors, and prevent costly manufacturing downtime.

Manufacturers, nowadays, are investing in wireless technologies to allow the engineers to acquire and control the real-time data of wireless sensor/actuator networks of the factory at anytime, anywhere. Moreover, the adoption of multiple network technologies in a single environment is becoming common today. As shown in Fig. 10.8, real-time process control and maintenance systems are equipped with wireless sensors/actuator networks on the plant floor and can be integrated with the back end enterprise software as well as the internet web services.

Data can be entered or acquired while the alerts/alarms can be notified through SMS or emails to the engineers at offices or remote locations. Incorporating short-range communication technologies like ZigBee/Bluetooth into the automation system will enable the engineers to collect and control real time sensors/actuators data from the factory floor, and internet-enabled handheld devices such as mobile phones/PDAs to connect with the outside world (Internet) make ubiquitous computing possible in industrial automation systems. Integration of such wide range of devices and different technologies together provide possible leverage strengths from one and another resulting in an efficient automated system.

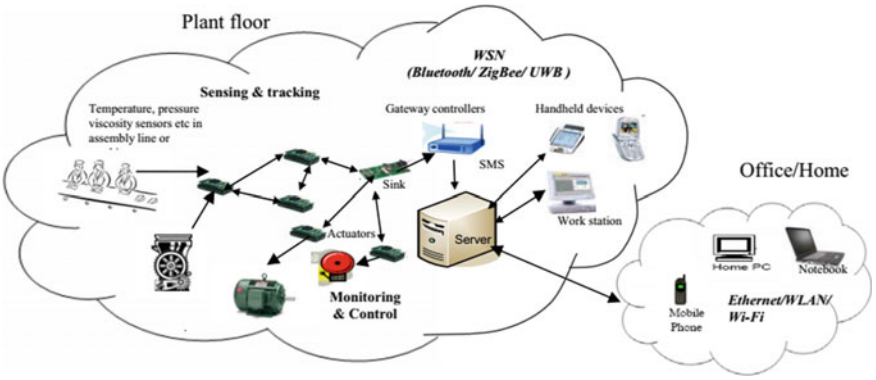


Fig. 10.8 Industrial networks

10.4.1 Industrial Mobile Robots

Autonomous mobile robots have been used widely in the industries. They are very suited for applications that are delicate, risky, heavy, and repetitive tasks. Most robots used in the industries are fixed and dedicated such as machine loading, welding, assembly/disassembly, etc. The use of intelligent mobile robots and wireless networks are being exploited in recent years. Unlike the traditional navigation that used limited embedded sensors and landmark objects to avoid obstacles in the paths, sensors, and processors are equipped on the robots to navigate based on information obtained from the networked sensors in the workspace. Such interactions among sensors in the field, sensors attached to the robots, laptop, and humans improve the environment perception of the workspace with collision-free navigation.

In [10], multi-robot task allocation is used where tasks are allocated explicitly to robots by a redeployed static sensor network. In this way, the robots can navigate efficiently throughout the workspace using the deployed sensor network to explore their environment. Potential applications include moving stacks of containers between machine rooms and warehouse, industrial floor cleaning robots, patrolling robots in unsupervised areas, or toxic waste cleanup, etc. Mobile robots can also be used to calibrate, recalibrate, or deliver power to sensors as an overall caretaker/service robot of the wireless sensor network. Mobile robots equipped with calibrated sensors will visit the field to collect data from many sensors and decide whether the sensors need to recalibrate.

10.4.2 Real-Time Inventory Management

Inventory management systems based on manual processes may cause out-of-stocks, expedited shipments, production slowdowns, excess buffer inventory, and billing

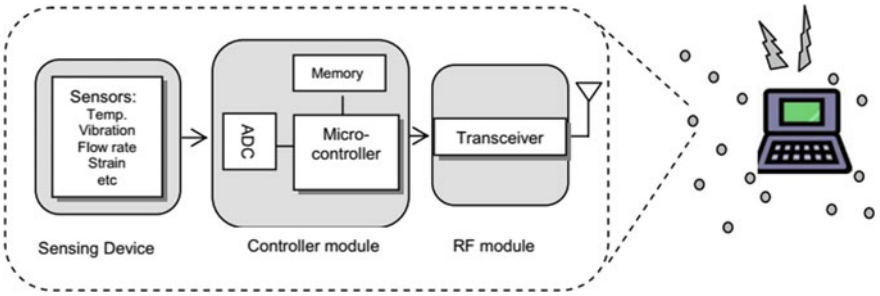


Fig. 10.9 Condition monitoring of machine

delays situations. With WSN technology, the inventory and asset could be monitored in real time and information such as the arrival of the raw materials, etc., could be routed across a distant to the gateway for management decision and control.

General Motors is implementing a real-time inventory tracking system [11]. The tracking process of the inventory is starting from the components suppliers, to the assembled cars in the factory, the car dealers and until the car buyers. Real time inventory tracking with WSN improves the visibility of materials, its location and asset utilization, reduces theft, provides the ability to immediately find equipment and ultimately increases supply chain efficiencies.

10.4.3 Process and Equipment Monitoring

WSNs can be used to enable remote monitoring of the health of machinery without the infrastructure needs for cabling. By continuously monitoring the temperature, pressure, vibrations, and power usage, etc., as shown in Fig. 10.9, the manufacturers can reduce unnecessary cost incurred due to failures or malfunctions in machinery.

Intel's EcoSense project group [12] is deploying a pilot preventive maintenance application that uses wireless sensor network to monitor the health of semiconductor fabrication equipment. With this advance real time monitoring system, the vibration signature of water purification equipment is analyzed and the data is used for preventive maintenance operations. In another trial deployment, Intel and BP has also collaborated and successfully experimented WSN on a crude oil tanker at Scotland to monitor machinery vibration to support preventive maintenance. The test has shown that the WSN can function well in a hostile shipboard environment where it would have to withstand temperature extremes, substantial vibration, and significant RF noise.

According to the United States Department of Energy, the electric motors in the industries consume 23% of energy in the U.S. industry. To counter such immense energy use, the Department has commissioned a project [13] to develop a low cost self-configuring wireless sensor network to apply monitoring and diagnostic systems

for use in electric motors. The sensors will have embedded intelligence and could measure parameters such as voltage, current, and temperature. This wireless transmitted information would then be used by the network's energy management system for the plant control. It is targeted to have a 5% energy reduction when implemented successfully.

10.4.4 Environment Monitoring

Wireless sensing can be used to provide solutions in the industry for leakage detection, climate reporting, radiation check, intrusion notification, etc. Emergency alerts could be sent to the operating managers requesting immediate preventive actions. With WSNs, the presence or the movement of abnormalities such as toxic chemical, biological, radioactive agent or unauthorized personnel can be tracked throughout the facility.

Leakage of flammable liquids and gases such as ammonia, chlorine, etc., in the petrochemical plants can cause heavy loss, risks for public, and hazardous emissions to the environment. Many oil and gas companies are now pilot testing WSNs and plan to deploy widely in near future. It is estimated that 2 million units could be deployed within the next 5 years [14]. With the wireless sensors, once leakage is detected, responsive actions can be activated such as emergency alert to the operators, automatic switch off of the machinery or trigger the sprinkler alarms, etc.

Running wires in the hazardous environments are not practical and also prohibitively expensive. To perform critical functions in sensitive and harsh environments such as propulsion test [15], WSNs are used by NASA to provide the remote sensing and monitoring of the engine testing with high reliability and accuracy. The parameters that are monitored include acoustic, strain, vibration level, vacuum level, temperature, etc.

In [16], a wireless sensor network based system of autonomous temperature logging of fish catches has been successfully tested off the Irish coast. The objective is to improve food safety and also the chain traceability. For the implementation, the individual fish box is equipped with a sensor node that transmits time stamped temperature data and node identification regularly to the base station on the ship. The processed information is then transmitted to the computer server on shore in real time via GSM modem.

10.5 Conclusions

Wireless sensor networks provide the industries with real-time tracking and remote monitoring/control capability on systems or devices. For some large factories, there is a need for the central office to track and log equipment location and status on the production floors. It is also required to track employees and visitors, usually with an

accuracy of better than 1 m. Current solutions employ WiFi for data communications and a separate system for locating equipment and personnel. UWB system can solve both communication and locating needs, possibly as part of a hierarchical system in which the UWB locating and communication system is installed for each floor and connected via WiFi for the whole factory. This chapter presented several industrial applications that take advantage of the wireless sensor network. It highlighted the wireless technologies and their standardization trend for process automation is explained. Introducing wireless technologies into plants and factories will reduce production cost and increase productivities more and more.

References

1. Quang PTA, Kim D-S (2015) Clustering algorithm of hierarchical structures in large-scale wireless sensor and actuators networks. J Commun Netw. ISSN: 1976–5541
2. Galloway B, Hancke G (2013) Introduction to industrial control networks. IEEE Commun Surveys Tutorials 15(2):860–880
3. Alliance Z (2008) Zigbee specification. Zigbee Document 053474r13, Zigbee alliance, May 2008
4. I. W. W. Group (2008) Draft standard ISA100. 11a. International Society of Automation, May 2008
5. Dinh NQ, Kim D-S (2012) Performance evaluation of priority CSMA-CA mechanism on ISA100.11a wireless network. Comput Stand Interface 34(1):117–123. ISSN: 0920-5489
6. Chen MND, Mok A (2010) WirelessHART: real-time mesh network for industrial automation, 1st edn. Springer Publishing Company, Incorporated
7. Zheng L (2010) Industrial wireless sensor networks and standardizations: the trend of wireless sensor networks for process automation. In: Proceedings of SICE annual conference, Aug 2010, pp 1187–1190
8. Akyildiz I, Kasimoglu I (2004) A protocol suite for wireless sensor and actor networks. In: IEEE radio and wireless conference, Sept 2004, pp 11–14
9. Hayashi H, Hasegawa T, Demachi K (2009) Wireless technology for process automation. In: ICCAS-SICE, Aug 2009, pp 4591–4594
10. Batalin MA, Sukhatme GS (2004) Using a sensor network for distributed multi-robot task allocation. In: Proceedings IEEE international conference on robotics and automation, vol 1, 158–164
11. Hochmuth P (2005) GM cuts the cords to cut costs. Mobility wirel Art Techworld
12. Technology@Intel Magazine (2005) Expanding usage models for wireless sensor networks, Aug 2005, pp 4–5
13. US Department of Energy (2004) Sensors and automation: Eaton wireless sensor network for advanced energy management solutions, June 2004
14. Kevan T (2005) Wireless sensors. Sensors Magazine online, Aug 2005
15. Solano WM, Junell J, Schmalzel JL, Shummard KC (2004) Implementation of wireless and intelligent sensor technologies in the propulsion test environment. In: Proceedings IEEE conference on sensor for industry, Jan 2004, pp 135–138
16. Crowley K, Frisby J, Edwards S, Murphy S, Roantree M, Diamond D (2004) Wireless temperature logging technology for the fishing industry. In: Proceedings IEEE sensors, Oct 2004, pp 571–574

Chapter 11
MAC Protocols for Energy-Efficient Wireless Sensor Networks

11.1 Introduction

Improvements in hardware technology have resulted in low-cost sensor nodes, which are composed of a single chip embedded with memory, a processor, and a transceiver. Low power capacities lead to limited coverage and communication range for sensor nodes compared to other mobile devices. Hence, for example, in target tracking and border surveillance applications, sensor networks must include a large number of nodes in order to cover the target area successfully.

Unlike other wireless networks, it is generally difficult or impractical to charge/replace exhausted batteries. That is why the primary objective in wireless sensor networks design is maximizing node/network lifetime, leaving the other performance metrics as secondary objectives. Since the communication of sensor nodes will be more energy consuming than their computation, it is a primary concern to minimize communication while achieving the desired network operation.

However, the medium-access decision within a dense network composed of nodes with low duty cycles is a challenging problem that must be solved in an energy-efficient manner. Keeping this in mind, we first emphasize the peculiar features of sensor networks, including reasons for potential energy waste at medium-access communication. Then, we give brief definitions for the key medium access control (MAC) protocols proposed for sensor networks, listing their advantages and disadvantages. Moreover, protocols that propose the integration of MAC layer with other layers are also investigated. Finally, the survey of MAC protocols is concluded with a comparison of investigated protocols and future directions are provided for researchers with regard to open issues that have not been studied thoroughly.

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_11

11.2 MAC Layer-Related Sensor Network Properties

Maximizing the network lifetime is a common objective of sensor network research, since sensor nodes are assumed to be dead when they are out of battery. Under these circumstances, the proposed MAC protocol must be energy efficient by reducing the potential energy wastes presented below. The types of communication patterns that are observed in sensor network applications should be investigated, since these patterns determine the behavior of the sensor network traffic that has to be handled by a given MAC protocol. The categorization of possible communication patterns is outlined, and the necessary MAC protocol properties suitable for a sensor network environment are presented.

11.2.1 Reasons of Energy Waste

When a node receives more than one packet at the same time, these packets are termed collided, even when they coincide only partially. All packets that cause the *collision* have to be discarded and retransmissions of these packets are required, which increase the energy consumption. Although some packets could be recovered by a *capture* effect, a number of requirements have to be achieved for successful recovery. The second reason for energy waste is *overhearing*, meaning that a node receives packets that are destined to other nodes. The third energy waste occurs as a result of *control packet overhead*. A minimal number of control packets should be used to make a data transmission. One of the major sources of energy waste is *idle listening*, that is, listening to an idle channel in order to receive possible traffic. The last reason for energy waste is *overemitting*, which is caused by the transmission of a message when the destination node is not ready. Given the above facts, a correctly designed MAC protocol should prevent these energy wastes.

11.2.2 Communication Patterns

Kulkarni [1] defines three types of communication patterns in wireless sensor networks: *broadcast*, *convergecast*, and *local gossip*. A broadcast pattern is generally used by a base station (sink) to transmit some information to all the sensor nodes of the network. Broadcasted information may include queries of sensor query processing architectures, program updates for sensor nodes, or control packets for the whole system. The broadcast communication pattern should not be confused with broadcast packets. For the former, all nodes of the network are intended receivers, whereas for the latter, the intended receivers are the nodes within the communication range of the transmitting node.

In some scenarios, the sensors that detect an event communicate with each other locally. This kind of communication pattern is called local gossip, where a sensor sends a message to its neighboring nodes within a range. After the sensors detect an event, they need to send what they perceive to the information center. That communication pattern is called convergecast, in which a group of sensors communicate to a specific sensor. The destination node could be a cluster head, a data fusion center, or a base station. In protocols that include clustering, cluster heads communicate with their members and thus the intended receivers may not be all neighbors of the cluster head, but just a subset of the neighbors.

11.2.3 Properties of a Well-Defined MAC Protocol

To design a good MAC protocol for wireless sensor networks, the following attributes must be considered [2]. The first attribute is energy efficiency. Hence, energy-efficient protocols are defined in order to prolong the network lifetime. Other important attributes are scalability and adaptability to changes. Changes in network size, node density, and topology should be handled rapidly and effectively for successful adaptation. Some of the reasons behind these network property changes are limited node lifetime, addition of new nodes to the network, and varying interference, which may alter the connectivity and hence the network topology. A good MAC protocol should gracefully accommodate such network changes. Other important attributes such as latency, throughput, and bandwidth utilization may be secondary in sensor networks. Contrary to other wireless networks, fairness among sensor nodes is not usually a design goal, since all sensor nodes share a common task [3].

11.3 Multiple-Access Consideration in Sensor Network Properties

Sensor networks consider communication needs of a collection of wireless devices, and not just the design of a single radio link. The algorithms and protocols that network devices is used to efficiently communicate are the topic of this section.

In a wireless network, the manner in which devices access and use the transmission medium (in this case, a wireless channel) is termed multiple access; within IEEE 802 terminology, it falls under the scope of the Multiple-Access Control (MAC) sublayer. All devices on the network must share the wireless channel since wireless communication is inherently a broadcast communications scheme and signals sent by one transmitter are heard at multiple locations. Thus, a major goal of the MAC is to limit/minimize the interference within the network. There are several well-known methods by which wireless devices can share a channel. These typically involve

transmitting signals that are orthogonal in one or more dimension such as time, frequency, or code.

11.3.1 Network Topologies

For further discussion about multiple accesses, we refer the reader to Fig. 11.1, which depicts a simple star network consisting of six nodes. Using IEEE 802.15.4 terminology, this collection of nodes is termed a PAN; and it is assumed to span a small (10 m) geographical area. Additionally, there are two types of nodes defined in the standard; an FFD and an RFD.

From the PAN control and multiple-access point of view, an FFD contains the software that enables PAN initiation, network formation, and control of the wireless channel for multiple accesses among the RFDs.

An FFD is commonly referred to as a "coordinator" due to its ability to provide the above functions. In the figure, the FFD node is depicted in the center of the PAN while the RFD nodes are shown surrounding the coordinator. The arrows indicate that the RFD devices are logically associated with the coordinator and rely on it for multiple-access services and data transport.

Figure 11.2 shows another example of a sensor network topology, typically referred to as a tree network. In this figure, we again consider both FFD and RFD devices as in Fig. 11.1.

The tree network can be viewed as an amalgamation of star networks (depicted by the dashed circles) where the star networks are connected together by linking

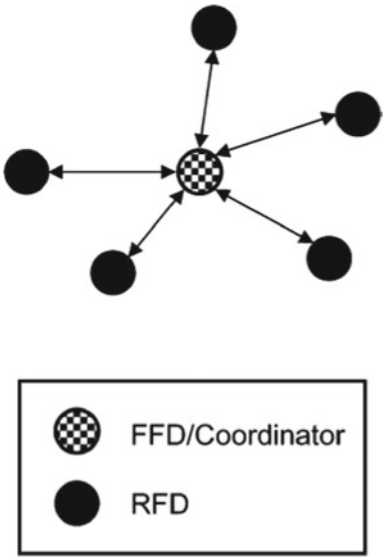


Fig. 11.1 A simple sensor network with a star topology

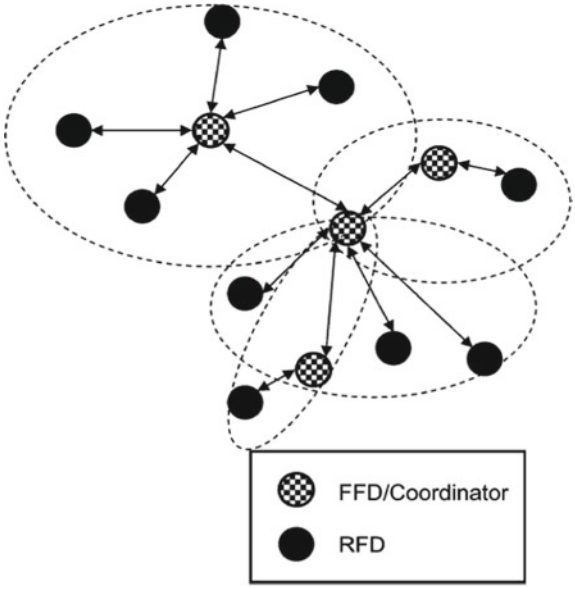


Fig. 11.2 Sensor network with a tree topology

the FFDs in each star together. Note here that data may need to be routed through multiple hops if devices want to communicate outside of their local star network. A third topology to consider is a mesh topology, which is similar to the multi-hop tree topology but with the addition of multiple links among the devices. (In a tree network, there exists only one path between any two devices.) The mesh topology in Fig. 11.3 provides reliability to the network in the form of redundant paths among the devices so, in the event of device or link failure, data may be rerouted.

When considering multiple-access methods, it is useful to understand how the topology effects the multiple-access requirements. Typically, a simple topology leads to simple multiple-access designs since there are fewer devices accessing the channel and thus less possibility of interference among the devices. More importantly, simple topologies can offer the ability to control access at a central point; such is the case of the star network where a single FFD device controls the timing of transmissions. More complex topologies require more careful planning of the channel access in order to minimize interference, but they do allow coverage of larger areas by a single network even with severely constrained transmit power, as is the case for UWB networks.

Given the topologies described above, we are now ready to discuss various multiple-access techniques. First, let us distinguish between two broad categories of multiple-access techniques: centralized and decentralized. In a centralized access scheme, a single node or small subset of nodes is responsible for controlling the transmissions of other devices in the network. In a decentralized scheme, each node is responsible for deciding if and when to transmit on the channel. Typically centralized schemes offer better efficiency and reliability since collisions can be more

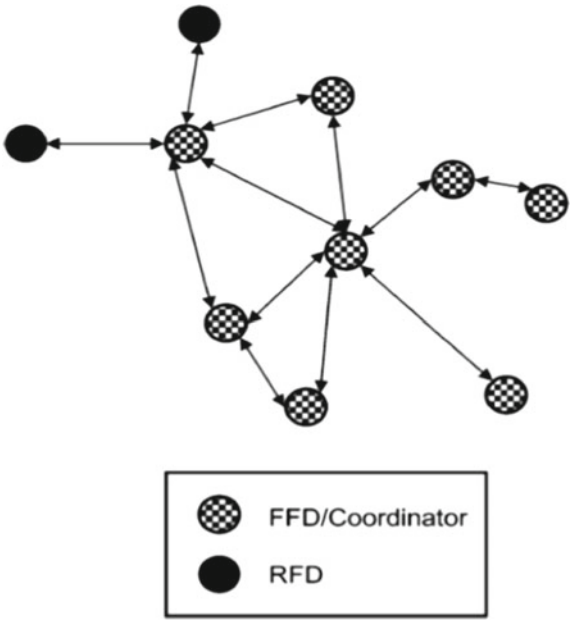


Fig. 11.3 Sensor network with a mesh topology

easily avoided, but this comes at the cost of increased complexity in the nodes that control the access as well as a need for network-wide information regarding the communication needs of every node in the network. Decentralized schemes tend to be simpler than centralized ones but less reliable due to the lack of network-wide knowledge and strong control, so that nodes have a higher probability of accessing the channel during other transmissions and thus causing interference to one another.

Distributed schemes are typically realized via handshaking-based approaches. Handshaking may prevent collisions, but note that additional messages for handshaking need to be transmitted. A device starts a request to send/clear to send (RTS/CTS) exchange on a common channel with its destination. If the channel is available, the subsequent data transmission uses a particular time hopping sequence proposed in the CTS. The reader is referred to for a detailed survey on medium-access control in ultra-wide-band wireless networks.

11.3.2 Time-Division Multiple Access (TDMA)

TDMA is a centralized scheme in which only one device transmits at any given time interval. We have essentially signals that are orthogonal in time; this is achieved by dividing the time axis into discrete no overlapping transmission intervals and assigning intervals to particular network devices. The devices then only transmit during their assigned time, and at all other times may listen to the channel to hear

transmissions from other devices. For the purpose of a sensor network, TDMA in this strict definition is not necessarily feasible. This is due to the fact that in order to fully coordinate the timing of transmissions from multiple devices, a global time reference is needed, i.e., the network would need to be synchronized. For a small network consisting of a few devices all within communication range, synchronization is possible. However, in many scenarios envisioned for sensor networks, network-wide synchronization and thus TDMA was not considered. Another issue with TDMA relates to the scheduling of packet transmissions among the nodes. In order for a controlling node to assign slots efficiently, it must have information regarding the amount of data each network node wishes to transmit. Several techniques have been developed to deliver such information to the controlling node. A simple approach is for the coordinator to poll each device to ascertain its current traffic load, and then, it may adjust the length of subsequent TDMA slots accordingly. However, when only a subset of nodes have data to send, the exchange of polling messages is wasteful of network bandwidth. This is generally the case with TDMA systems where there is a tradeoff between the amount of scheduling efficiency that can be achieved and the amount of control information that must be passed among the FFD and RFDs.

11.3.3 Carrier-Sense Multiple Access (CSMA) and ALOHA

CSMA can be viewed as a distributed version of TDMA. In this scheme, each node in the network attempts to avoid colliding with other transmissions. The basic idea is that each node senses the wireless channel prior to transmitting a packet to determine if the channel is in use. If the channel is idle, the node can then transmit its packet; otherwise, the node waits a for a time period of random length and repeats the sensing and transmission. Thus CSMA attempts to arrange transmissions in orthogonal time intervals. The advantage of a CSMA scheme over TDMA is that it is distributed. Additionally, each node will attempt to access the channel only when it has data ready for transmission. This eliminates the need for complex scheduling. However, CSMA suffers from some well-known problems. First and foremost is the "hidden terminal" problem in which a node that senses the channel may not be within radio range of all nodes in the network.

Thus, even though a node may determine that the channel is idle and transmit, communication may be taking place elsewhere in the network. These transmissions have the potential to interfere. Additionally, CSMA relies on the ability of performing an accurate channel sensing. This seemingly simple operation can be quite difficult in UWB-TH-IR systems. This difficulty arises from the fact that UWB transmission is extremely low power and require knowledge of the spreading code for effective dispreading. Thus a node would ideally check all possible spreading codes before declaring an idle channel. In large networks using many codes, this may not be feasible.

11.3.4 Frequency-Division Multiple Access (FDMA)

Analogously to TDMA, FDMA assigns orthogonal frequency channels to various devices. This can be achieved by dividing the frequency spectrum into no overlapping segments and assigning these segments to individual devices for their transmissions. Within the context of UWB systems, this multiple-access technique has several problems. First, regulatory requirements require that UWB devices transmit signals with a bandwidth no smaller than 500 MHz. Thus in order to support N users, the system bandwidth would need to be at least 500_N MHz. So we see that in order to support multiple simultaneous users, each device must be able to receive and process extremely wideband signals. Secondly, depending on the duplexing method, network-wide synchronization may still be needed. This is the case when considering half duplex communication where devices may be either transmitting or receiving. In this case, the system must schedule which devices are to be transmitting and which are to be receiving during each time instant. This type of scheduling is difficult to achieve without some form of global time reference. Additionally, scheduling broadcast or multicast traffic becomes problematic in FDMA networks with half duplex devices. Full duplex devices mitigate the scheduling problem somewhat, but these are intrinsically more costly, as full duplex system require essentially two radios per device, and each radio would need to operate over a large system bandwidth. Still, usage of different frequency bands allows a very good separation of signals that would be difficult to separate, e.g., by CDMA. For the above reason, FDMA is useful, e.g., to separate closely spaced networks, and is used for this purpose also in IEEE 802.15.4a.

11.3.5 Code-Division Multiple Access (CDMA)

CDMA assigns (quasi-) orthogonal spreading codes to individual devices, which then multiply their symbol stream by the assigned code. In its most general form, CDMA encompasses all the spreading schemes. Receivers can differentiate among different devices by correlating the received signal with each user's assigned code. CDMA networks do not have the scheduling issues associated with TDMA and FDMA techniques described above. Since they rely on signal processing at the receiver to separate transmissions from multiple users, CDMA allows the simultaneous transmissions (in time and/or frequency). CDMA is also attractive for UWB sensor networks because the spreading factor in a UWB system is so large, theoretically, many simultaneous transmission can be supported.

The IEEE 802.15.4a standard relies on this large spreading factor and the ability to resolve multiple users to enable reuse of frequency bands. That is, multiple networks may be deployed within a single frequency band. Thus, every device on the network need only listen for packets that contain the correct code and then can synchronize its receivers to decode the subsequent data.

11.4 Proposed MAC Layer Protocols

In this section, a wide range of MAC protocols defined for sensor networks are described briefly by stating the essential behavior of the protocols wherever possible. Moreover, the advantages and disadvantages of these protocols are presented.

11.4.1 Sensor-MAC

Locally managed synchronizations and periodic sleep-listen schedules based on these synchronizations form the basic idea behind the Sensor-MAC (S-MAC) protocol [2]. Neighboring nodes form virtual clusters so as to set up a common sleep schedule. If two neighboring nodes reside in two different virtual clusters, they wake up at the listen periods of both clusters. A drawback of the S-MAC algorithm is this possibility of following two different schedules, which results in more energy consumption via idle listening and overhearing.

Schedule exchanges are accomplished by periodic SYNC packet broadcasts to immediate neighbors. The period for each node to send a SYNC packet is called the synchronization period. Figure 11.4 represents a sample sender–receiver communication. Collision avoidance is achieved by a carrier sense (represented as CS in the figure).

Furthermore, RTS/CTS packet exchanges are used for unicast-type data packets. S-MAC also includes the concept of message passing, in which long messages are divided into frames and sent in a burst. With this technique, one may achieve energy savings by minimizing communication overhead at the expense of unfairness in medium access.

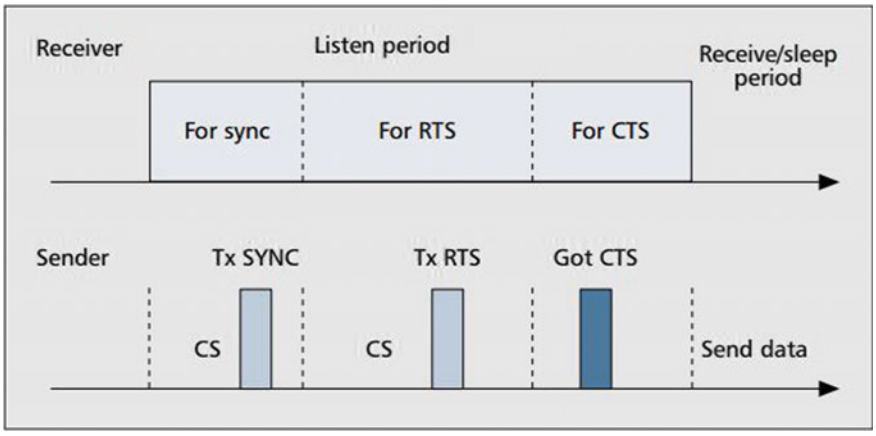


Fig. 11.4 The S-MAC messaging scenario

Periodic sleep may result in high latency, especially for multi-hop routing algorithms, since all intermediate nodes have their own sleep schedules. The latency caused by periodic sleeping is called sleep delay. The adaptive listening technique is proposed to improve the sleep delay and thus the overall latency. In that technique, the node that overhears its neighbor's transmissions wakes up for a short time at the end of the transmission. Hence, if the node is the next-hop node, its neighbor could pass data immediately. The end of the transmissions is known by the duration field of the RTS/CTS packets.

Advantages—The energy waste caused by idle listening is reduced by sleep schedules. In addition to its implementation simplicity, time synchronization overhead may be prevented by sleep schedule announcements.

Disadvantages—Broadcast data packets do not use RTS/CTS, which leads to increasing collision probability. Adaptive listening incurs overhearing or idle listening if the packet is not destined to the listening node. Sleep and listen periods are predefined and constant, which decreases the efficiency of the algorithm under variable traffic load.

11.4.2 WiseMAC

Hoiydi [4] proposed the "Spatial TDMA and CSMA with Preamble Sampling" protocol in which all sensor nodes is defined to have two communication channels. The data channel is accessed using TDMA, whereas the control channel is accessed by CSMA. The WiseMAC [5] protocol is similar to Hoiydi's work but requires only a single-channel. WiseMAC protocol uses nonpersistent CSMA (np-CSMA) with preamble sampling as in [4] to decrease idle listening. In the preamble sampling technique, a preamble precedes each data packet for alerting the receiving node. All nodes in a network sample the medium with a common period, but their relative schedule offsets are independent. If a node finds the medium busy after it wakes up and samples the medium, it continues to listen until it receives a data packet or the medium becomes idle again. The size of the preamble is initially set to be equal to the sampling period. However, the receiver may not be ready at the end of the preamble, due to factors such as interference, which causes the possibility of over emitting-type energy waste. Moreover, over emitting is increased with the length of the preamble and the data packet, since no handshake is done with the intended receiver.

To reduce the power consumption incurred by the predetermined fixed-length preamble, WiseMAC offers a method to dynamically determine the length of the preamble. That method uses the knowledge of the sleep schedules of the transmitter node's neighbors. The nodes learn and refresh their neighbor's sleep schedule during every data exchange as part of the Acknowledgment message. In that way, every node keeps a table of the sleep schedules of its neighbors. Based on the neighbors' sleep schedule tables, WiseMAC schedules transmissions so that the destination node's sampling time corresponds to the middle of the sender's preamble. To decrease the

possibility of collisions caused by that specific start time of a wake up preamble, a random wake-up preamble is advised.

Another parameter affecting the choice of the wake-up preamble length is the potential clock drift between the source and the destination. A lower bound for the preamble length is calculated as the minimum of destination's sampling period, *Tw*, and the potential clock drift with the destination, which is a multiple of the time since the last ACK packet arrived. Considering this lower bound, a preamble length (*Tp*) is chosen randomly. Figure 11.5 presents the WiseMAC concept.

Advantages—The simulation results show that WiseMAC performs better than one of the S-MAC variants [5]. Besides, its dynamic preamble length adjustment results in better performance under variable traffic conditions. In addition, clock drifts are handled in the protocol definition, which mitigates the external time synchronization requirement.

Disadvantages—The main drawback of WiseMAC is that decentralized sleep-listen scheduling results in different sleep and wake-up times for each neighbor of a node. This is an important problem especially for broadcast-type communication, since broadcasted packets will be buffered for neighbors in sleep mode and delivered many times as each neighbor wakes up. However, this redundant transmission will result in higher latency and power consumption. In addition, the hidden-terminal problem accompanies the WiseMAC model, as in the Spatial TDMA and the CSMA with Preamble Sampling algorithm. That is because WiseMAC is also based on nonpersistent CSMA. This problem will result in collisions when one node starts to transmit the preamble to a node that is already receiving another node's transmission where the preamble sender is not within range.

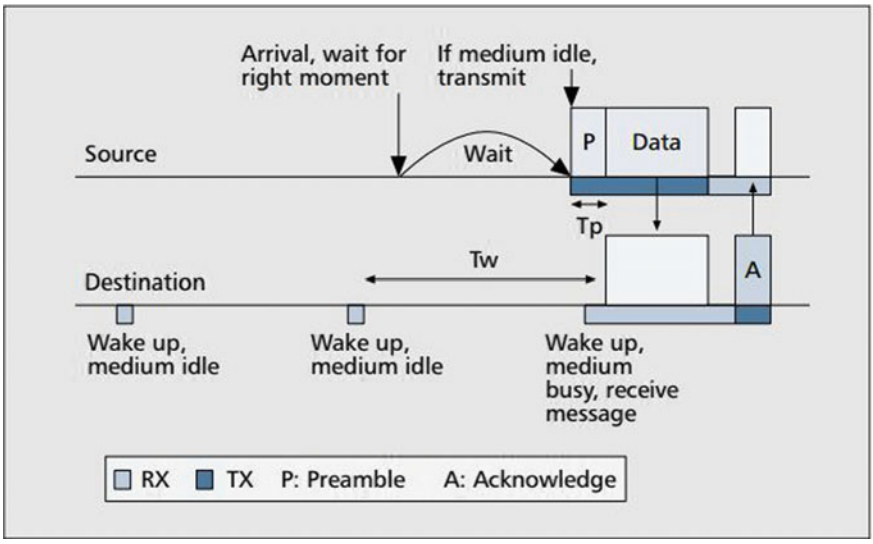


Fig. 11.5 The WiseMAC concept

11.4.3 Traffic-Adaptive MAC Protocol

TRAMA [5] is a TDMA-based algorithm proposed to increase the utilization of classical TDMA in an energy-efficient manner. It is similar to Node Activation Multiple Access (NAMA) [5], in which for each time slot a distributed election algorithm is used to select one transmitter within each two-hop neighborhood. This kind of election eliminates the hidden-terminal problem and hence ensures that all nodes in the one-hop neighborhood of the transmitter will receive data without any collision. However, NAMA is not energy efficient and incurs overhearing.

Time is divided into random-access and scheduled-access (transmission) periods. The random-access period is used to establish two-hop topology information and the channel access is contention-based within that period. A basic assumption is that, with the information passed by the application layer, the MAC layer can calculate the transmission duration needed, which is denoted as *SCHEDULE_INTERVAL*. Then, at time t, the node calculates the number of slots for which it will have the highest priority among two-hop neighbors within the period $[t, t+SCHEDULE\_INTERVAL]$. The node announces the slots it will use as well as the intended receivers for these slots with a schedule packet. Additionally, the node announces the slots for which it has the highest priority but it will not use. The schedule packet indicates the intended receivers using a bitmap whose length is equal to the number of its neighbors. Bits correspond to one-hop neighbors ordered by their identities. Since the receivers of those messages have the exact list and identities of the one-hop neighbors, they find out the intended receiver. When the vacant slots are announced, potential senders are evaluated for reuse of those slots. Priority of a node on a slot is calculated with a hash function of node's and slot's identities.

Analytical models for the delay performances of TRAMA and NAMA protocols are also presented and supported by simulations [6]. Delays are found to be higher, as compared to those of contention-based protocols, due to a higher percentage of sleep times.

Advantages—Higher percentage of sleep time and less collision probability are achieved, as compared to CSMA-based protocols. Since the intended receivers are indicated by a bitmap, less communication is performed for the multicast and broadcast types of communication patterns, compared to other protocols.

Disadvantages—Transmission slots are set to be seven times longer than the random-access period. However, all nodes are defined to be either in receive or transmit states during the random-access period for schedule exchanges. This means that without considering the transmissions and receptions, the duty cycle is at least 12.5%, which is a considerably high value. For a time slot, every node calculates each of its two-hop neighbors' priorities on that slot. In addition, this calculation is repeated for each time slot, since the parameters of the calculation change with time.

11.4.4 Sift

Sift [7] is a MAC protocol proposed for event-driven sensor network environments. The motivation behind Sift is that when an event is sensed, the first R of N potential reports are the most crucial part of messaging and have to be relayed with low latency. Jamieson et al. use a nonuniform probability distribution function of picking a slot within the slotted contention window. If no node starts to transmit in the first slot of the window, then each node increases its transmission probability exponentially for the next slot, assuming that the number of competing nodes is small.

In [7], Sift was compared with the 802.11 MAC protocol and it was shown that Sift decreases latency considerably when there are many nodes trying to send a report. Since Sift is a contention slot assignment algorithm, it is proposed to coexist with other MAC protocols like S-MAC. Based on the same idea, CSMA/p^* [8] is proposed where p^* is a nonuniform probability distribution that optimally minimizes latency. However, Tay et al. state that the probability distribution function of Sift to pick a slot is approximate to CSMA/p^*.

Advantages—Very low latency is achieved for many traffic sources. Energy consumption is traded-off for latency, as indicated below. However, when the latency is an important parameter of the system, slightly increased energy consumption must be accepted. The Sift algorithm could be tuned to incur less energy consumption. High energy consumption is a result of the arguments indicated below.

Disadvantages—One of the main drawbacks is increased idle listening caused by listening to all slots before sending. The second drawback is increased overhearing. When there is an ongoing transmission, nodes must listen until the end in order to contend for the next transmission, which causes overhearing. Besides, system-wide time synchronization is needed for slotted contention windows. That is why the implementation complexity of Sift would be larger than protocols not utilizing time synchronization.

11.4.5 DMAC

Converge cast is the most frequent communication pattern observed within sensor networks. Unidirectional paths from sources to the sink could be represented as data-gathering trees. The principal aim of DMAC [9] is to achieve very low latency for converge cast communications, but still be energy efficient.

DMAC could be summarized as an improved Slotted Aloha algorithm in which slots are assigned to the sets of nodes based on a data-gathering tree, as shown in Fig. 11.6. Hence, during the receive period of a node, all of its child nodes have transmit periods and contend for the medium. Low latency is achieved by assigning subsequent slots to the nodes that are successive in the data transmission path.

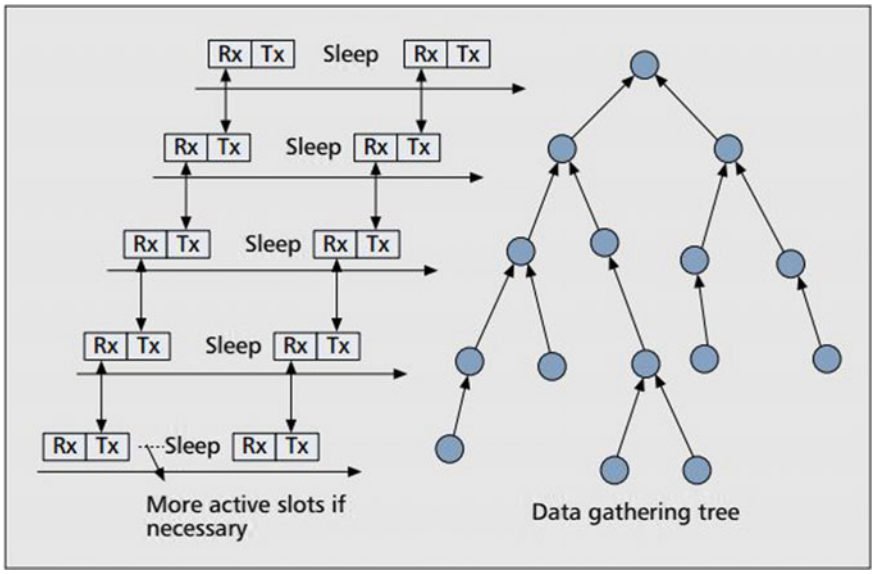


Fig. 11.6 A data-gathering tree and its DMAC implementation

Advantages—DMAC achieves very good latency compared to other sleep/listen period assignment methods. The latency of the network is crucial for certain scenarios, in which DMAC could be a strong candidate.

Disadvantages—Collision avoidance methods are not utilized; hence, when a number of nodes that have the same schedule (the same level in the tree) try to send to the same node, collisions will occur. This is a possible scenario in event-triggered sensor networks. Besides, the data transmission paths may not be known in advance, which precludes the formation of the data-gathering tree.

11.4.6 Timeout-MAC/Dynamic Sensor-MAC

The static sleep-listen periods of S-MAC result in high latency and lower throughput, as indicated above. Timeout-MAC (T-MAC) [10] is proposed to enhance the poor results of the S-MAC protocol under variable traffic loads. In T-MAC, the listen period ends when no activation event has occurred for a time threshold TA. The decision for TA is presented along with some solutions to the early sleeping problem defined in [10]. Variable loads in sensor networks are expected, since the nodes that are closer to the sink must relay more traffic and traffic may change over time. Although T-MAC gives better results under these variable loads, the synchronization of the listen periods within virtual clusters is broken. This is one of the reasons for the early sleeping problem.

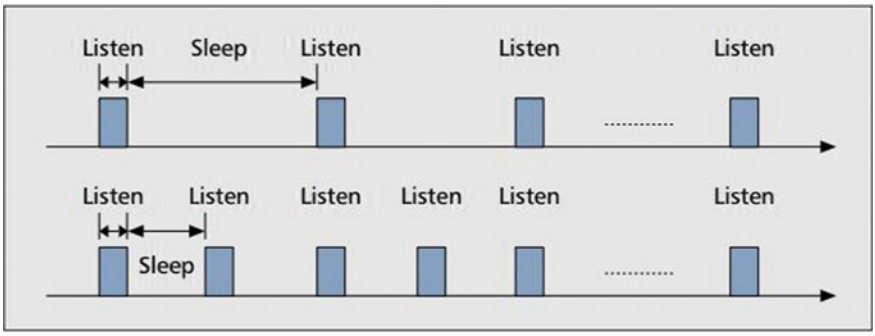


Fig. 11.7 DSMAC duty cycle doubling

Dynamic Sensor-MAC (DSMAC) [11] adds a dynamic duty cycle feature to S-MAC. The aim is to decrease the latency for delay-sensitive applications. Within the SYNC period, all nodes share their one-hop latency values (the time between the reception of a packet into the queue and its transmission). All nodes start with the same duty cycle. Figure 11.7 conceptually depicts DSMAC duty-cycle doubling. When a receiver node notices that the average one-hop latency value is high, it decides to shorten its sleep time and announces it within the SYNC period. Accordingly, after a sender node receives this sleep-period decrement signal, it checks its queue for packets destined to that receiver node. If there is one, it decides to double its duty cycle when its battery level is above a specified threshold. The duty cycle is doubled so that the schedules of the neighbors will not be affected. The latency observed with DSMAC is better than that observed with S-MAC. Moreover, it is also shown to have better average power consumption per packet.

11.4.7 Integration of MAC with Other Layers

Limited research has been carried out on integrating different network layers into one layer or to benefit from cross-layer interactions between routing and MAC layers for sensor networks. For instance, Safwat et al. proposed two routing algorithms that favor the information about successful/unsuccessful CTS or ACK reception [12].

Cui et al. looked at MAC/physical layer integration and Routing/MAC/physical layer integration [13]. They proposed a variable length TDMA scheme in which the slot length is assigned according to some criteria for optimum energy consumption in the network. Among these criteria, the most crucial ones are information about the traffic generated by each node and the distances between each node pair. Based on these values, they formulated a Linear Programming (LP) problem in which the decision variables are normalized time-slot lengths between nodes. They solve this LP problem using an LP solver that returns the optimum number of time slots for each node pair as well as the related routing decisions for the system. The proposed

solution could be beneficial in scenarios where the required data would be prepared. However, it is generally difficult to have the node-distance information and the traffic generated by the nodes. Besides, the LP solver can only be run on a powerful node. The dynamic behavior of sensor networks will require online decisions which are very costly to calculate and hard to adapt to an existing system.

Multi-hop Infrastructure Network Architecture (MINA) is another method for integrating MAC and routing protocols [14]. Ding et al. proposed a layered multi-hop network architecture in which the network nodes with the same hop-count to the base station are grouped into the same layer. Channel access is a TDMA-based MAC protocol combined with CDMA or FDMA. The super-frame is composed of a control packet, a beacon frame, and a data transmission frame. The beacon and data frames are time slotted. In the clustered network architecture, all members of a cluster submit their transmission requests in beacon slots. Accordingly, the cluster head announces the schedule of the data frame.

The routing protocol is a simple multi-hop protocol where each node has a forwarder node at one nearer layer to the base station. The forwarding node was chosen from candidates based on the residual energies. Ding et al. then formulated the channel allocation problem as an NP-complete problem and proposed a suboptimal solution. Moreover, the transmission range of the sensor nodes is a decision variable, since it affects the layering of the network (the hop-counts change). Simulations were run to find a good range of values for a specific scenario. The proposed system in [14] is a well-defined MAC/Routing system. However, the tuning of the range parameter is an important task that should be done at system initialization. In addition, all node-to-sink paths are defined at the startup and are defined to be static, since channel frequency assignments of nodes are done at the startup accordingly. This makes the system intolerant to failures.

Geographic Random Forwarding (GeRaF) is actually proposed as a routing protocol, but the underlying MAC algorithm is also defined in the work, which is based on CSMA/CA [15]. This work gives a complete (but not integrated) solution for a sensor network's communication layers. The difficulty of the system proposed is its need for an additional radio, which is used for the busy-tone announcement. Rugin et al. [16] and Zorzi [15] improved GeRaF by reducing it to a one-channel system. However, the sensor nodes' and their neighbors' location information is needed for those protocols. Besides, the forwarding node is chosen among nodes that are awake at the time of the transmission request. That may result in routing with more power consumption and an increase in latency.

11.5 Open Issues and Conclusion

Figure 11.8 gives a comparison of the MAC protocols investigated. The column heading "Time Synchronization Needed" indicates whether the protocol assumes that the time synchronization is achieved externally and "Adaptivity to Changes" indicates the ability to handle topology changes. The two S-MAC variants, namely,

	Time sync needed	Comm. pattern support	Type	Adaptivity to changes
S-MAC/T-MAC/DSMAC	No	All	CSMA	Good
WiseMAC	No	All	np-CSMA	Good
TRAMA	Yes	All	TDMA/CSMA	Good
Sift	No	All	CSMA/CA	Good
DMAC	Yes	Convergecast	TDMA/Slotted Aloha	Weak

Fig. 11.8 Comparison of MAC protocols

T-MAC and DSMAC, have the same features as S-MAC (Fig. 11.4). The cross-layer protocols include additional layers other than the MAC layer and are not considered in this comparison. Although there are various MAC layer protocols proposed for sensor networks, there is no protocol accepted as a standard. One of the reasons for this is that the MAC protocol choice will, in general, be application dependent, which means that there will not be one standard MAC for sensor networks. Another reason is the lack of standardization at lower layers (physical layer) and the (physical) sensor hardware.

TDMA has a natural advantage of collision-free medium access. However, it includes clock drift problems and decreased throughput at low traffic loads due to idle slots. The difficulties with TDMA systems are synchronization of the nodes and adaptation to topology changes when these changes are caused by insertion of new nodes, exhaustion of battery capacities, broken links due to interference, the sleep schedules of relay nodes, and scheduling caused by clustering algorithms. The slot assignments, therefore, should be done with regard to such possibilities. However, it is not easy to change the slot assignment within a decentralized environment for traditional TDMA, since all nodes must agree on the slot assignments.

In accordance with common networking lore, CSMA methods have a lower delay and promising throughput potential at lower traffic loads, which generally happens to be the case in wireless sensor networks. However, additional collision avoidance or collision detection methods should be employed. FDMA is another scheme that offers a collision-free medium, but it requires additional circuitry to dynamically communicate with different radio channels. This increases the cost of the sensor nodes, which is contrary to the objective of sensor network systems.

CDMA also offers a collision-free medium, but its high computational requirement is a major obstacle for the less energy consumption objective of sensor networks. In pursuit of low computational cost for wireless CDMA sensor networks, there has been limited effort to investigate source and modulation schemes, particularly signature waveforms, designing simple receiver models, and other signal synchronization problems. If it is shown that the high computational complexity of CDMA could be traded-off against its collision-avoidance feature, CDMA protocols could also be considered as candidate solutions for sensor networks. Lack of comparisons of

TDMA, CSMA, or other medium-access protocols in a common framework is a crucial deficiency of the literature.

Common wireless networking experience also suggests that link-level performance alone may provide misleading conclusions about the system performance. A similar conclusion can be drawn for the upper layers as well. Hence, the more layers contributing to the decision, the more efficient the system can be. For instance, the routing path could be chosen depending on the collision information from the medium-access layer. Moreover, layering of the network protocols creates overheads for each layer, which causes more energy consumption for each packet. Therefore, integration of the layers is also a promising research area that needs to be studied more extensively.

References

1. Kulkarni S (2004) TDMA service for sensor networks. In: 24th international conference on distributed computing systems workshops, Mar 2004, pp 604–609
2. Ye W, Heidemann J, Estrin D (2004) Medium access control with coordinated adaptive sleeping for wireless sensor networks. IEEE/ACM Trans Netw 12(3):493–506
3. Quang PTA, Kim D-S (2015) Clustering algorithm of hierarchical structures in large-scale wireless sensor and actuators networks. J Commun Netw. IF: 0.747, ISSN: 1976-5541
4. El-Hoiydi A (2002) Spatial TDMA and CSMA with preamble sampling for low power ad hoc wireless sensor networks. In: Seventh international symposium on computers and communications, pp 685–692
5. Xie M, Wang X (2008) An energy-efficient TDMA protocol for clustered wireless sensor networks. In: CCCM'08. ISECS international colloquium on computing, communication, control, and management, vol 2, Aug 2008, pp 547–551
6. Zheng D, Ge W, Zhang J (2009) Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach. IEEE Trans Inf Theory 55(1):205–222
7. Jamieson K, Balakrishnan H, Tay YC (2003) SIFS: a MAC protocol for event-driven wireless sensor networks. Technical report
8. Tay YC, Jamieson K, Balakrishnan H (2004) Collision-minimizing CSMA and its applications to wireless sensor networks. IEEE J Sel Areas Commun 22(6):1048–1057
9. Lu G, Krishnamachari B, Raghavendra C (2004) An adaptive energy-efficient and low-latency mac for data gathering in wireless sensor networks. In: Proceedings of 18th international parallel and distributed processing symposium, Apr 2004
10. van Dam T, Langendoen K (2003) An adaptive energy-efficient mac protocol for wireless sensor networks. In: Proceedings of the 1st international conference on embedded networked sensor systems, series. SenSys'03, pp 171–180
11. Lin P, Qiao C, Wang X (2004) Medium access control with a dynamic duty cycle for sensor networks. In: IEEE wireless communications and networking conference, vol 3, Mar 2004, pp 1534–1539
12. Safwat A, Hassanein H, Mouftah H (2003) ECPS and E2LA: new paradigms for energy efficiency in wireless ad hoc and sensor networks. In: IEEE global telecommunications conference GLOBECOM'03, vol 6, Dec 2003, pp 3547–3552
13. Cui S, Madan R, Goldsmith A, Lall S (2005) Joint routing, MAC and link layer optimization in sensor networks with energy constraints. In: IEEE international conference on communications, ICC 2005, vol 2, May 2005, pp 725–729
14. Ding J, Sivalingam K, Kashyapa R, Chuan LJ (2003) A multi-layered architecture and protocols for large-scale wireless sensor networks. In: IEEE 58th vehicular technology conference, vol 3, Oct 2003, pp 1443–1447

15. Zorzi M (2004) A new contention-based mac protocol for geographic forwarding in ad hoc and sensor networks. In: IEEE international conference on communications, vol 6, June 2004, pp 3481–3485
16. Rugin R, Mazzini G (2004) A simple and efficient MAC-routing integrated algorithm for sensor network. In: IEEE international conference on communications, vol 6, June 2004, pp 3499–3503

Chapter 12
Cooperative Multi-channel Access for Industrial Wireless Networks Based 802.11 Standard

12.1 Introduction

Cooperative communication, which can achieve spatial diversity by exploiting distributed virtual antennas of cooperative nodes, has attracted much attention because of its ability to mitigate fading in wireless networks. Previous studies have shown that significant gain can be obtained through cooperative communication in terms of reliability, coverage range, and energy efficiency.

IEEE 802.11 networking is entering a new phase with the ongoing standardization of IEEE 802.11s and the recent introduction of Wi-Fi Direct technology. The new technologies will allow 802.11 devices to easily communicate directly with each other by forming a mesh network, which will open up new avenues for device-to-device communication. The full potential of these avenues lies in allowing multiple simultaneous transmissions in a given radio neighborhood, which is a challenge for current 802.11 mesh networks based on a fixed single-channel communication architecture. To address this, multi-channel communication techniques are being investigated in order to allow multi-channel access in 802.11 mesh. This chapter focuses on the practical utilization of a novel cooperation approach at the Media Access Control (MAC) layer to allow multi-channel access in 802.11 mesh networks.

CAMMAC-802.11 and Distributed Interference-Aware Relay Selection (DIRS) algorithm have been proposed for IEEE 802.11-based wireless networks with multiple source-destination pairs. CAMMAC-802.11 is the new protocol that employs a novel approach to control plane cooperation called Distributed Information Sharing (DISH), which was introduced to solve the Multi-Channel Coordination (MCC) problem. Real-world experiments were carried out in a mesh testbed to evaluate the protocol performance and compare it with that of the IEEE 802.11 MAC. CAMMAC-802.11 is a feasible protocol choice for the 802.11 mesh networks and can indeed reap the benefits of using multiple channels [1]. Besides, to mitigate the impact of inter-node interference, [1] proposed a DIRS algorithm that selects a relay node with consideration of both: inter-node interference and channel conditions. The algorithm

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_12

is efficient as shown by simulation results and it is also practical as it only requires local information instead of network topology information. Moreover, in order to decrease end-to-end delay and AP access delay, a cooperative association mechanism has been proposed for managing wireless mesh networks [2].

12.2 Throughput Enhancement

Throughput maximization is a key challenge to numerous applications in Wireless Mesh Networks (WMNs). As a potential solution, cooperative communications, which may increase link capacity by exploiting spatial diversity, has attracted a lot of attention [3, 4]. However, if link scheduling is considered, this transmission mode may perform worse than direct transmission in terms of end-to-end throughput, since the sending/receiving of cooperative relays incurs extra interferences. CAMMAC-802.11 was used in [5] as a method to maintain throughput at a high level compared with using typical 802.11 standard, and [6] used directional antennas while the authors in [7] introduced a special algorithm called NEgotiation-based Throughput Maximization Algorithm (NETMA) to maximize this metric.

12.2.1 CAMMAC-802.11

CAMMAC-802.11 uses a control channel to allow nodes to (1) negotiate for data channels, (2) alert nodes of MCC problem and (3) share neighbor information. The control channel can be from a licensed spectrum or an unlicensed band. If no licensed band is available and channels in the unlicensed bands are occupied, CAMMAC-802.11 network still can use the least occupied channel as the control channel, but will have to share the bandwidth with other 802.11 networks present on that channel. CAMMAC-802.11 protocol uses the 802.11 (a/b/g/n) physical layer (PHY), so the standard channel bandwidth, both for the control channel and data channel, will be PHY based. For the purpose of cooperation, nodes maintain a spectrum usage table and a neighbor table. The spectrum usage table stores channels that are currently in use in the network. Each entry consists of Tx Rx MAC addresses, data channel, and corresponding expiration time. The neighbor table contains information about node's neighbors and non-conflicting neighbors, i.e., neighbors that cannot hear each other. The CAMMAC-802.11 is specifically designed as a multi-channel extension of the IEEE 802.11 MAC. The original CAMMAC [8] cannot be used because its design is generic and not compatible with the 802.11 MAC. The design and protocol implementation of CAMMAC-802.11 was indicated in [5].

Four experiments were conducted, one based on the 802.11 MAC which used a single channel and the others based on CAMMAC-802.11 with 1 control and 4(CAM 4chnl)/3(CAM 3chnl)/2(CAM 2chnl) data channels. CAM 4chnl experiment was free of channel conflict as enough channels were available for 4 pairs and we used

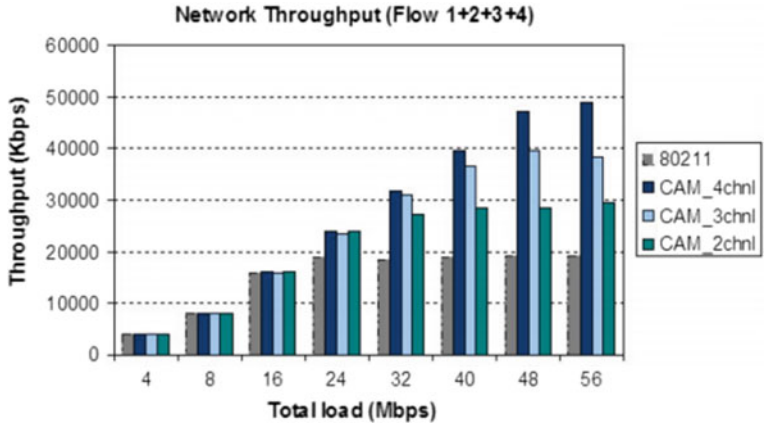


Fig. 12.1 802.11 MAC versus CAMMAC-802.11 throughput performance

a proper channel selection strategy (choose the previously used channel first). For CAM 3chnl and CAM 2chnl experiments, conflict may occur within the network due to lack of enough data channels. In the experiments, we used train size of 20 (TxOp (transmission opportunity) = 11 ms) for CAMMAC-802.11. The measured aggregated throughput (Flow 1 + 2+3 + 4) result is shown in Fig. 12.1. As shown, the performance of the CAMMAC-802.11 slightly trails behind that of 802.11 when the traffic load (sum of applied loads at the 4 Tx) is low (16 Mbps or less). This is because for low-traffic loads when a single channel is sufficient for channel contention in the case of 802.11, the control session of the CAMMAC-802.11 acts as an overhead and degrades the performance.

In Fig. 12.2, paper [5] has presented the per node throughput results with the varying network size (number of Tx–Rx pairs) and 14 Mbps traffic load per transmitter. The CAMMAC-802.11 results are for varying number of data channels. For a given network size, we start experiment with the number of data channels equal to the number of pairs and keep repeating the experiment with one less data channel till we have only two data channels left. For all the 802.11 experiments, we use only one data channel. The CAMMAC_xless legend in Fig. 12.2 represents an experiment, where the number of data channels used is x less than the number of pairs in the experiment. For example, CAMMAC legend (or CAMMAC_0less) represents the experiment with the number of data channels equal to the number of pairs. If we compare 802.11 throughput for 2 pairs with that of CAMMAC-802.11 throughput in which 4 pairs with 3 data channels (CAMMAC_1less) and CAMMAC-802.11 compared with 4 pairs with 2 data channels (CAMMAC_2less), we find that 802.11 throughput falls in between and is comparable with the CAMMAC_2less throughput. Keeping in mind that the hardware CAMMAC-802.11 will perform better, the comparison can roughly be interpreted as: if CAMMAC-802.11 has n channels, it can support n times the number of nodes supported by the 802.11 standard with comparable per node throughput. This shows the advantage of multi-channel access, when it comes to the scalability of mesh networks.

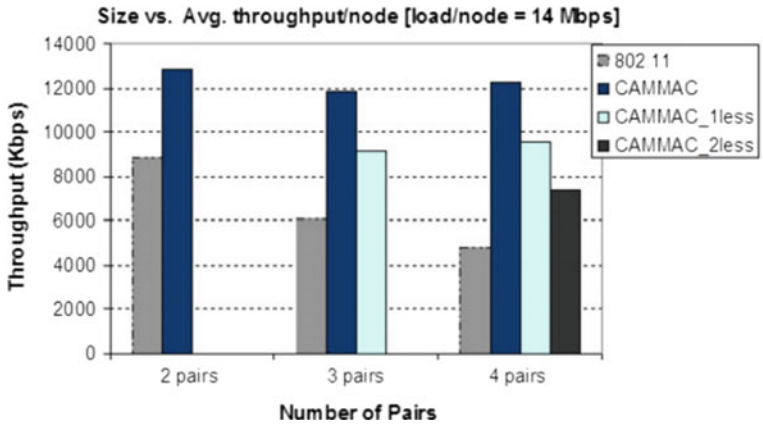


Fig. 12.2 Network size versus avg. throughput/node

12.2.2 Using Directional Antennas

Paper [6] has studied the throughput maximization problem in cooperative WMNs under multiple constraints (i.e., antenna mode selection, transmission mode selection, link scheduling, and flow routing). Considering the special features of directional antennas and cooperative communications, it first extended the links and classified them into cooperative links/general links. Then, depending on different beamforming strategies at the transmitters, it formed cooperative conflict graphs to describe the interference relationship among those extended links. After that, it mathematically formulated the end-to-end throughput maximization problem with the fairness of radio resource allocation. By numerical simulations, we demonstrate that the scheme, in which the transmitters beamform to the receivers and cooperative communications is considered, is better than the other schemes in terms of end-to-end throughput in cooperative WMNs.

In Fig. 12.3, we compare the throughput performance of different schemes with different combinations of beamforming strategies and transmission modes in multi-hop cooperative WMNs. In this figure, both DA + CC and OA + CC denote the cross-layer designs with a joint consideration of antenna mode selection and transmission mode selection, where the transmitter uses directional antennas and beamform to the receiver in DA + CC, and the transmitter uses omnidirectional antennas in OA + CC; DA and OA denote the schemes only considering antenna mode selection, where directional antennas and omnidirectional antennas are employed by the transmitters in the network, respectively; OA-All denotes the scheme without any consideration of antenna mode selection or transmission mode selection, where all nodes use omni-directional antennas for transmissions. As shown in Fig. 12.3, DA + CC outperforms the other schemes in terms of end-to-end throughput in cooperative WMNs. It is not surprising that OA-All has the worst performance because it only considers traditional link scheduling and flow routing, and has no concern about either transmission

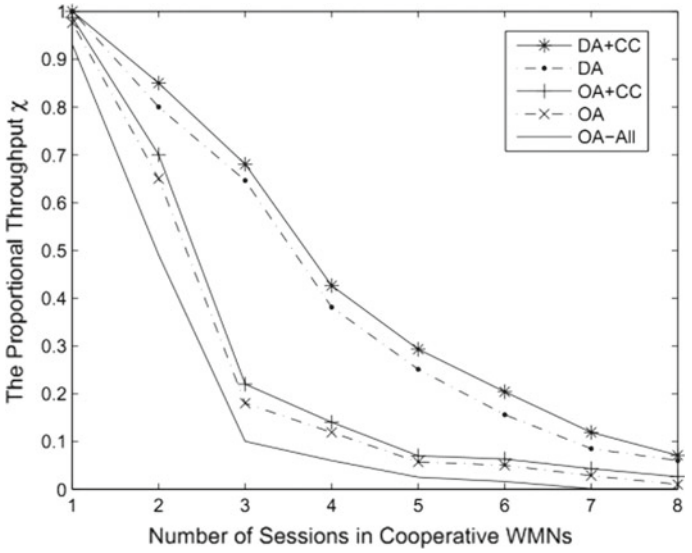


Fig. 12.3 Performance comparison of different schemes with different antenna modes and transmission modes in multi-hop cooperative WMNs: 20 nodes

mode selection or antenna mode selection. OA is better than OA-All in the sense that OA allows the receivers to beamform to the transmitters. OA is inferior to OA + CC since OA ignores the opportunities of cooperative communications. Although DA neglects the cooperative communications as well, it is still superior to OA + CC due to the trunking throughput gain brought by directional antennas of the transmitters. Compared with DA and OA + CC, DA + CC further improves the end-to-end throughput in cooperative WMNs, even though DA + CC sacrifices the opportunities to use the potential cooperative relays beyond the beamforming area of the transmitters.

12.2.3 Negotiation-Based Throughput Maximization Algorithm

For the case of cooperative access points, the authors in [7] presented a NEgotiation-based Throughput Maximization Algorithm (NETMA), which adjusts the operating frequency and power level among access points autonomously, from a game theoretical perspective. The authors showed that this algorithm converges to the optimal frequency and power assignment which yields the maximum overall throughput with arbitrarily high probability. Moreover, they analyze the scenario where access points belong to different regulation entities and hence noncooperative.

NETMA converges to the optimum solution with arbitrarily high probability. For the noncooperative scenarios, the authors show the existence and the inefficiency

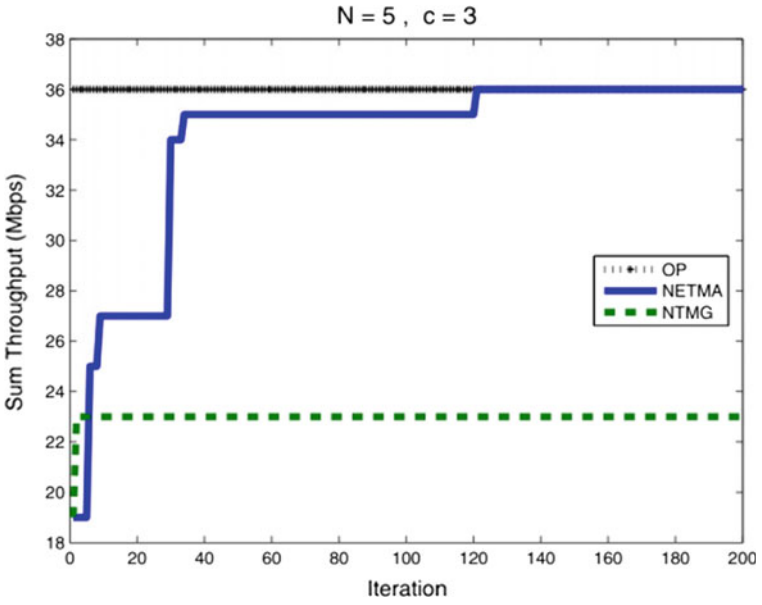


Fig. 12.4 Performance evaluation of the wireless mesh access network with $N = 5$ and $c = 3$

of Nash equilibria due to the selfish behaviors. To bridge the performance gap, we propose a linear pricing scheme which tremendously improves the performance in terms of overall throughput. The analytical results are verified by simulation.

As indicated by the OP curve, the global optimum obtained by enumeration approach functions as the upper bound of the overall throughput. In Fig. 12.4, we observe that NETMA gradually catches up with the global optimum as negotiations go. As expected, the noncooperative APs yield remarkably inferior performance in terms of overall throughput, depicted by the Noncooperative Throughput Maximizing Game (NTMG) curve. The inefficiency is due to the selfish behavior that APs transmit at the maximum power and are regardless of the interference. The existence of Nash equilibrium in both Cooperative Throughput Maximization Game (CTMG) and NTMG are substantiated by the convergence of curves in Fig. 12.4.

Figure 12.5 pictorially depicts the performance inefficiency of NTMG caused by the noncooperative APs, which transmit at the maximum power. The average throughput per AP is calculated by averaging the results of 50 simulations, for each value of the side length d. In Fig. 12.5, it is worth noting that as the side length d gets bigger, the performance gap between NETMA and NTMG reduces. The reason is that when the area is large, the impact of mutual interference is less severe and so is the performance deterioration. However, when the network is crowded, the selfish behaviors are remarkably devastating. The throughput improvement is illustrated as Noncooperative Throughput Maximization Game with Pricing (NTMGP) in Fig. 12.5. It is noticeable that by utilizing the proposed pricing scheme, the efficiency of Nash equilibrium is dramatically enhanced, especially for crowded networks. Therefore, the selfish incentives of the noncooperative APs have been effectively suppressed.

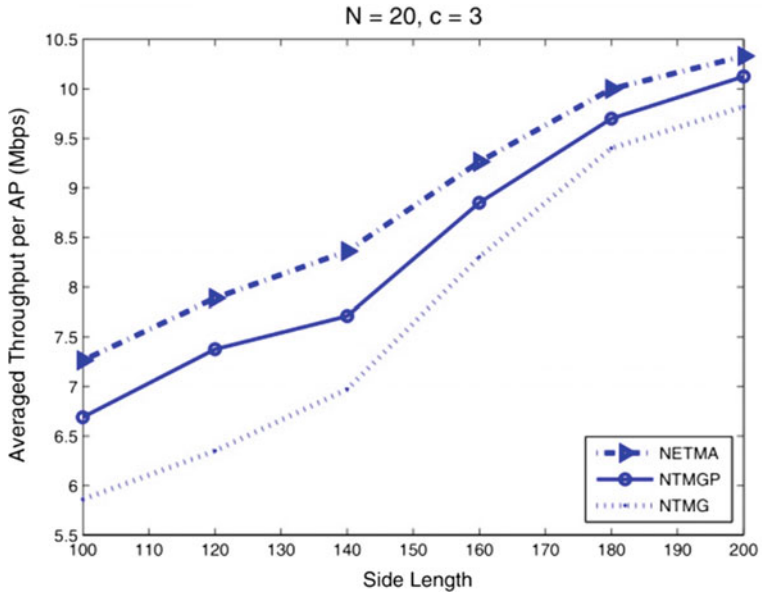


Fig. 12.5 Performance evaluation of the wireless mesh access network with $N = 20$ and $c = 3$

12.3 Access Delay

In [2], the concept of cooperative association was introduced, where the stations (STAs) can share useful information in order to improve the performance of the association/handoff procedures. The association/handoff procedures are important components in a balanced operation of 802.11-based wireless mesh networks. Furthermore, it introduces a load-balancing mechanism that can control the communication load of each mesh AP in a distributed manner. It has simulated a VoIP application in the 802.11-based wireless mesh network. In its simulations, we have uniformly placed several VoIP clients in the network. We run different simulation scenarios, where we vary the number of the VoIP sessions that are supported in parallel.

First of all, we measure the average local client access delay in the network. In practice, this delay reflects the time from when the packet is generated until it leaves the client interface. Figure 12.6 depicts the average VoIP client access delay. The load-balancing mechanism (enhanced with cooperation) achieves the lower client access delays in the network. The load-balancing mechanism minimizes the channel access delay, while it provides a cell breathing to the overloaded cells. The associated STAs are optimally associated in order to maintain a balanced network operation. Consequently, the load-balancing mechanism keeps the client access delay in low level while the traditional 802.11 operation overloads the network and the client access delay is continually increased. In high-load conditions, the delay improvement that is introduced by the load balancing mechanism is quite impressive.

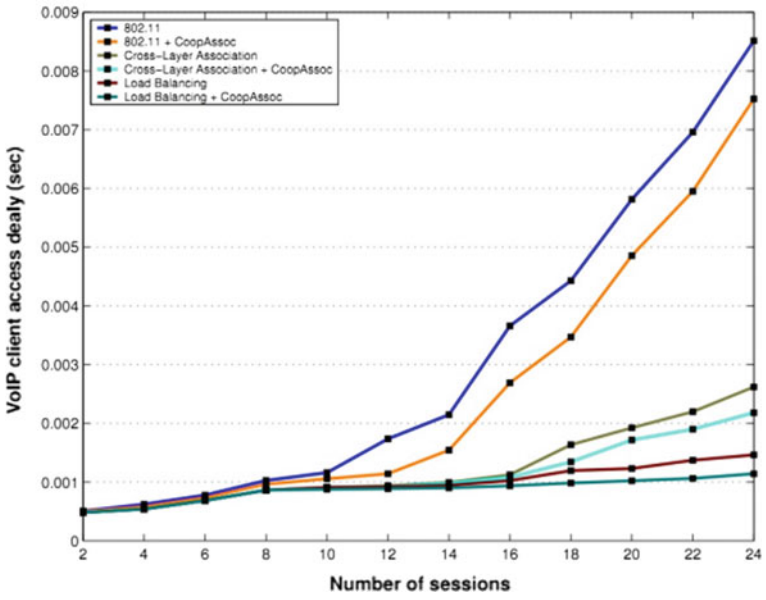


Fig. 12.6 Average VoIP delays

Figure 12.7 depicts the average local VoIP AP access delay in the network. This delay is the time between the arrival of a VoIP packet to the AP until it is either successfully transmitted over the wireless mesh network or dropped. It is clear that we get the same simulation results with the client access delay. The load-balancing mechanism (enhanced with cooperation) has the best performance. In 802.11, the overloaded APs (in high load conditions) have a lot of traffic to forward to the mesh backhaul network. The main consequence is that the VoIP packets have to wait for a long time to be transmitted by the APs, introducing in this way huge AP access delays.

In Fig. 12.8, we observe the average end-to-end delay in the VoIP packet transmission. The end-to-end delay is affected by the previous two kinds of delays that we have described in detail and the routing delay that is introduced in the backhaul network. The load-balancing mechanism (enhanced with cooperation) achieves lower end-to-end delays in the network. Especially in high-load network operation, the delay improvement is huge. This improvement is true due to the fast VoIP client and AP access in the network, the effective link aware AODV (Ad Hoc On-Demand Distance Vector) routing protocol in the mesh backhaul and the sophisticated cell breathing achieved by the load balancing mechanism in overloaded cells. The most interesting result is depicted in Fig. 12.8, the pure 802.11 operation can support at most 14 sessions in parallel while the proposed load-balancing mechanism has the capability to support 24 sessions in parallel. Therefore, we gain approximately 72.5% network performance improvement. The network capabilities are expanded by the use of the sophisticated load-balancing mechanism.

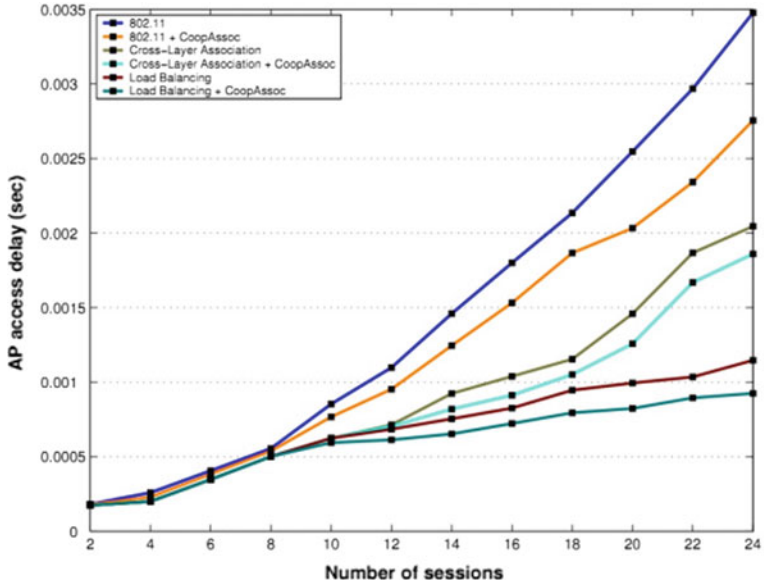


Fig. 12.7 Average AP access delay

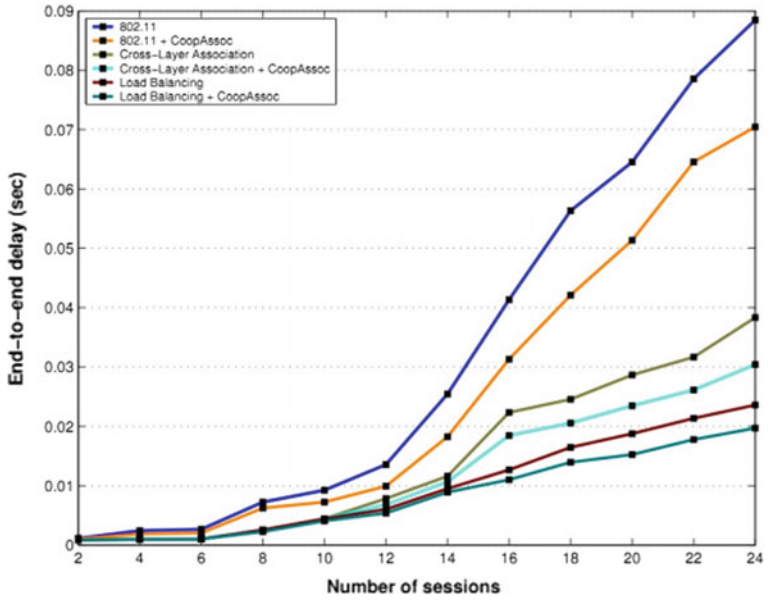


Fig. 12.8 Average end-to-end delay

12.4 Mitigating the Impact of Inter-node Interference

Paper [1] studies the interference-aware relay selection for IEEE 802.11 DCF-based wireless networks with multiple active source–destination pairs. It first illustrated that relay selection without considering inter-node interference will degrade rather than improve the network performance in some cases, and further propose an algorithm called Distributed Interference-Aware Relay Selection (DIRS) to select the relay node with considerations of both inter-node interference and channel conditions. Under DIRS algorithm, each source–destination pair only requires local information to select a relay node independently without any network topology knowledge. Through simulation, it demonstrated the effectiveness of the proposed DIRS algorithm and illustrate that inter-node interference can be mitigated by effectively selecting relays using DIRS.

Figure 12.9 reveals the total network throughput varying with source nodes traffic load that is the value of x. As each node traffic load increases, the total network throughput of the three approaches: our proposed relay transmission, direct transmission and traditional relay transmission all increase up to saturation, however, the proposed relay transmission always significantly outperforms direct transmission and traditional relay transmission. In Fig. 12.9, it is worth pointing out the fact that in the particular case of traditional relay transmission the total network throughput decreases significantly after reaching saturation before it finally stabilizes. This is because of the fact that the traditional relay transmission mode, exploits a relay selection algorithm based purely on wireless channel conditions and without the consideration of inter-node interference. It is neglecting the impact of inter-node interference that translates into the aforementioned decrease in throughput that we can appreciate in the graph and which will greatly affect network performance of cooperative networks with high network traffic. The proposed algorithm (DIRS) overcomes this by accounting for inter-node interference when performing relay transmission.

Figure 12.10 depicts the collision probability adopting the three different transmission modes. The collision probability of the proposed relay transmission, direct transmission, and traditional relay transmission all stabilize as the source nodes traffic load increase, however, the consideration of inter-node interference by the proposed relay transmission results in a collision probability that is always lower than direct transmission and traditional relay transmission. The simulation studies in this section, on the one hand, demonstrate the effectiveness of the proposed DIRS algorithm and on the other hand, they also illustrate that the inter-node interference can be mitigated by effectively selecting relays considering both the wireless channel conditions and Channel Idle Ratio (CIR) in IEEE 802.11 distributed coordination function (DCF)-based wireless networks with multiple source–destination pairs.

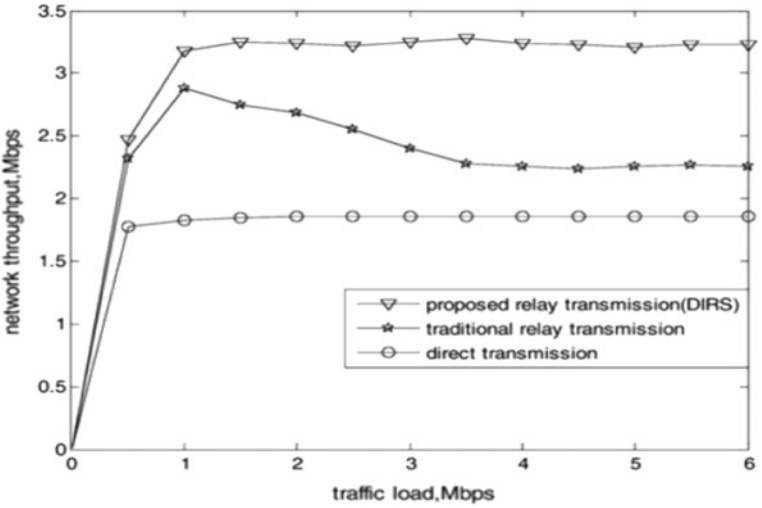


Fig. 12.9 Network throughput comparison of different transmission modes

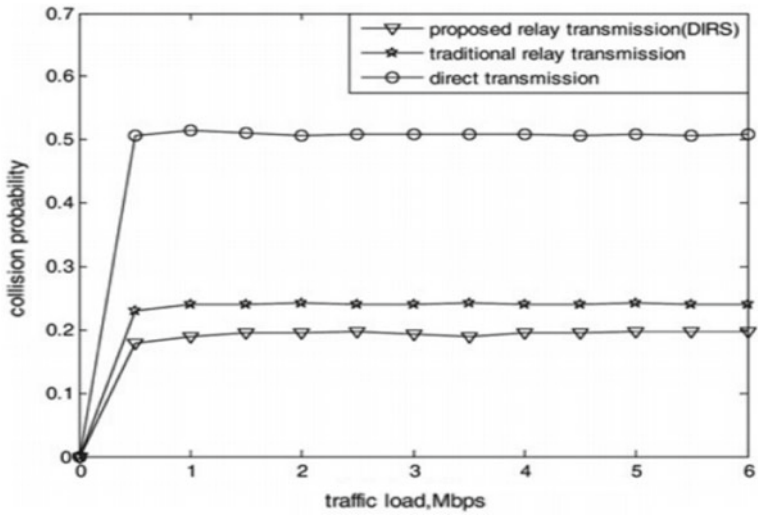


Fig. 12.10 Collision probability adopting three different transmission modes

12.5 Conclusions

In this chapter, by using CAMMAC-802.11 the throughput of industrial wireless mesh networks can be enhanced. We can also achieve this advantage by using directional antennas and negotiation-based throughput maximization algorithm. In order to eliminate the association/reassociation delays, a new association mechanism which

introduces cooperation between the STAs in a wireless mesh network has been proposed. Another aspect of this chapter is that inter-node interference can be mitigated by effectively selecting relays using DIRS. In future work, we will focus on the combination of methods to obtain much more effective results and consider the algorithms and protocols in overlapping channels and Linux open-source drivers.

References

1. Shi C, Zhao H, Garcia-Palacios E, Ma D, Wei J (2012) Distributed interference-aware relay selection for ieee 802.11 based cooperative networks. IET Netw 1(2):84–90
2. Athanasiou G (2012) Cooperative management of wireless mesh networks. In: Wireless Days (WD), 2012 IFIP, pp 1–6
3. Tan DD, Kim D-S (2015) Interference-aware relay assignment scheme for multi-hop wireless networks for wireless networks. Wirel Netw 1–13. IF: 1.055, ISSN: 1570-8705
4. Tan DD, Kim D-S (2014) Dynamic traffic-aware routing algorithm for multi-sink wireless sensor networks. Wirel Netw 20(6):1239–1250. IF: 1.055, ISSN: 1572-8196
5. Singh S, Motani M (2012) Cooperative multi-channel access for 802.11 mesh networks. IEEE J Sel Areas Commun 30(9):1684–1693
6. Pan M, Yue H, Li P, Fang Y (2012) Throughput maximization of cooperative wireless mesh networks using directional antennas. In: 1st IEEE international conference communications in China (ICCC), pp 10–14
7. Song Y, Zhang C, Fang Y (2007) Throughput maximization in multi-channel wireless mesh access networks. In: IEEE international conference on network protocols, pp 11–20
8. Luo T, Motani M, Srinivasan V (2009) Cooperative asynchronous multichannel mac: design, analysis, and implementation. IEEE Trans Mob Comput 8(3):338–352

Chapter 13
802.11 Medium Access Control DCF and PCF: Performance Comparison

13.1 Introduction

Nowadays, 802.11 standard [1] increasingly plays an important role in daily life as well as in the industry by various applications and advantages such as low cost, quick deployment, flexible configuration, user mobility support, etc. Thus, 802.11 also attracts much attention of researchers. Along with evolutions of this standard from 802.11 *b*, *g*, *n* to *ac,* the main difference is in modulation method and improvement at PHY layer, medium access control mechanisms almost not change except HCF which is an optional mechanism for delay-sensitive applications. An evolution in medium access control mechanisms is needed for higher performance 802.11 standard when the improvements at PHY layer are more close to theoretical limitation.

There are two different access mechanisms that are widely used to gain access to the shared wireless medium: the basic access mechanism, called the Distributed Coordinate Function (DCF), and a centrally controlled access mechanism, called the Point Coordinate Function (PCF). DCF implements the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) algorithm with binary exponential backoff algorithm for accessing the medium. This mechanism provides best-effort type of service for the transfer of data. DCF uses either two-way handshaking or four-way handshaking technique while transmitting the data. While two-way handshaking uses explicit acknowledgement (ACK) for receipt confirmation, RTS and CTS frames are used by four-way handshaking technique. In the PCF, access point within each Basic Service Set (BSS) network performs the role of the Point Coordinator (PC). Each superframe consists of a Contention Period (CP), where DCF is used and a Contention-Free Period (CFP) where PCF is used. Together the CFP and CP are repeated after every CFP Repetition Interval (CFPR). PCF must coexist with the DCF method. Hence PCF periods will occur periodically.

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_13

13.2 IEEE 802.11 Media Access Protocols

This section briefly covers the IEEE 802.11 medium access protocols. The 802.11 standard specifies a common Medium Access Control (MAC) Layer, which provides a variety of functions that support the operation of 802.11-based wireless LANs. The MAC Layer is responsible for coordinating the sharing of the physical layer (PHY). The physical layer standards like IEEE 802.11, 802.11a and 802.11b share the same MAC layer architecture. The 802.11 MAC layer uses the 802.11 PHY layer for carrier sensing, transmission, and receiving the 802.11 frames. The 802.11 MAC layer defines two modes of operation: Distributed Coordinate Function (DCF) and Point Coordinate Function (PCF) (Fig. 13.1).

In a wireless channel, transmissions are separated by inter-packet gaps known as Inter-Frame Spaces (IFS). Channel access is granted based on different priority classes. These classes are mapped on different gap durations. Distributed-IFS (DIFS, also known as DCF inter-frame space), Priority-IFS (PIFS, also known as PCF inter-frame spacing), and Short-IFS (SIFS) are the different time intervals being used in the wireless domain. While SIFS is the shortest time duration with the highest priority, DIFS is the longest time duration with the lowest priority. In addition, IEEE 802.11 MAC employs another technique based on virtual carrier-sensing mechanism to counter collisions. The technique uses a special entity known as Network Allocation Vector (NAV). The NAV value specifies a node, the amount of time that it has to wait before the channel will be available. The node transmitting the frames will include the remaining duration of transmission, which is copied on to the NAV value by the non-transmitting nodes. A node is not allowed to transmit when the NAV value is nonzero even in the case of the CSMA operation reveals an idle channel. These time frames and the NAV value are used in both DCF and PCF to provide collision-free transmission.

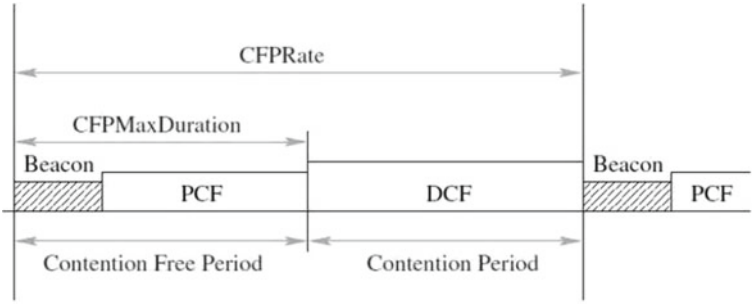


Fig. 13.1 IEEE 802.11 superframe DCF and PCF

13.2.1 Distributed Coordinate Function (DCF)

DCF employs carrier sense multiple access with collision avoidance (CSMA/CA) mechanism to access the channel. In addition, DCF also includes a binary exponential random backoff mechanism to ensure low collision probability. It is best known for its asynchronous data transmission (or best-effort service). DCF is the basic medium access mechanism for both ad hoc and infrastructure modes. In DCF mode, a station should ensure that the medium is free before it starts transmitting data over the channel. In the case of a busy channel, the node has to wait until the transmission by other node is complete. In addition, the node has to wait for another IFS amount of time to provide a sufficient gap between subsequent frame transmissions.

The nodes operating in DCF mode use the distributed inter-frame space (DIFS) (typically 50 ms for 802.11b) to transmit MAC frames. There are two basic rules that need to be followed by every node operating in the DCF mode. First every station that has data to transmit should confirm that the channel was idle for at least DIFS amount of time. Second, it should wait a random amount of time when it has another datagram to transmit after a successful transmission or it has a datagram to transmit but it was denied channel access, as the channel was busy. More specifically, the station selects a random number called backoff time, in the range of 0 and contention window (CW). Each time the carrier sense operation detects the medium to be idle, the backoff timer decrements the backoff time. If a collision occurs after the expiration of the backoff timer, the CW is doubled and the new backoff procedure will be initiated.

IEEE 802.11b standard specifies the transmission of an acknowledgement for every successful datagram receipt. This would enable the transmitter to detect the collision and retransmit the data. The transmitter expects an acknowledgement within SIFS time frame. Only after receiving an ACK frame correctly, the transmitter assumes that the data frame was delivered successfully. The SIFS and DIFS time-frames enable the communicating pairs to complete the frame exchange sequence without collision. Figure 13.2 describes the various phases involved in the DCF mode. It can be observed that the data transmission has to go through phases like DIFS deferral, contention phase, data transmission, SIFS deferral, and ACK transmission phases before completing the frame exchange.

While DCF is fair with all the nodes present in the network, it does not work well with the nodes transmitting/receiving multimedia traffic. This is because of the fact that every node has to contend for the media in which case a node with multimedia traffic may get very few chances to transmit data across the network. In [2, 3], the authors have concluded that even though DCF provides deadlock-free channel access, there are chances that a node may completely be starved of channel access.

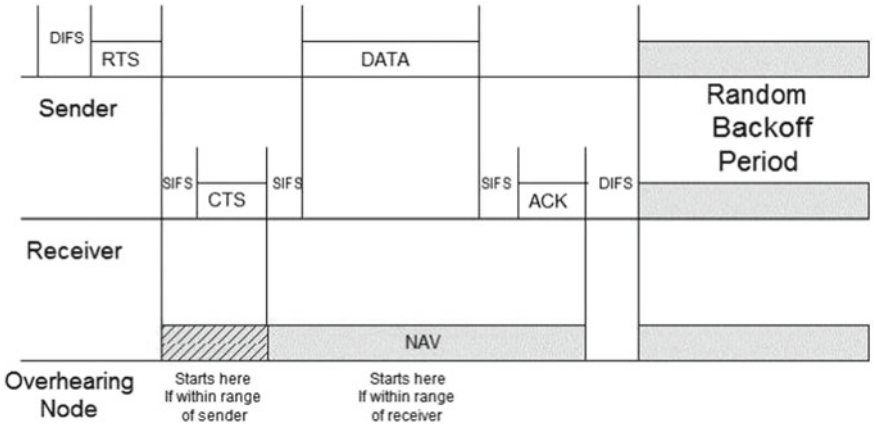


Fig. 13.2 Example for distributed coordination function

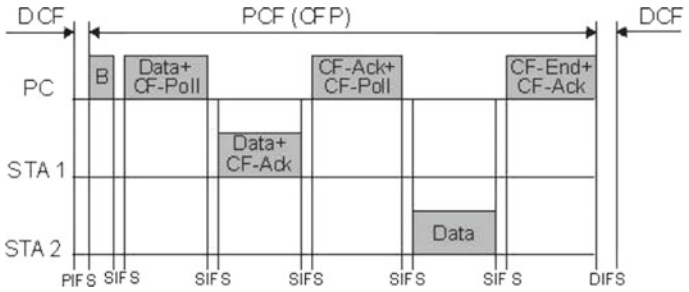


Fig. 13.3 Example for point coordination function

13.2.2 Point Coordinate Function (PCF)

The PCF is an optional capability that provides contention-free frame transfer. In a BSS, an Access Point (AP) also performs the functions of a Point Coordinator (PC). All stations have to obey the medium access rules of the PCF, because these rules are based on the DCF, and they set their NAV at the beginning of each Contention-Free Period (CFP). Figure 13.3 shows the working of PCF mechanism.

An active PC needs to be present with an AP, which restricts PCF operation to infrastructural networks. Data frames sent by, or in response to polling by, the PC during the CFP shall use the appropriate data subtypes. To start the CFP, PC gains the control of the medium by waiting PIFS period of time and then sends a beacon frame. It is a control frame that has all the attributes used in the CFP period. If PC has some data to send to the node then it sends data D1 along with a CF-POLL frame in order to poll the station. Only those stations that are PCF enabled respond to the CF-POLL. In this way, PC builds a polling list of all the PCF-enabled nodes. This list is organized according to the nodes' MAC addresses. Upon receiving CF-POLL,

stations wait for SIFS period of time and then either respond with data frame U1 along with the acknowledgement of the D1 data received or if they don't have any data to send then they will send a null frame in order to release the medium. This way PC will not wait for the PCF stations that don't have any data to send. After waiting for SIFS period, if the PC does not receive the data or null frame then it will repeat the polling of the same node. It keeps repeating this until a maximum number of poll failures (configurable parameter) occurs. It then drops that node in that CFP period and moves to the next station. That node is polled first in the subsequent CFP periods and the same process repeats in the subsequent CFP periods. The size of the CFP interval is also manageable. The CFP repetition interval is used to determine the frequency with which PCF occurs. The maximum size of the CFP will determine the CF-Max-Duration field that is manageable too.

The operating characteristics of the PCF are such that all stations are able to operate properly in the presence of a BSS in which a PC is operating, and, if associated with a point-coordinated BSS, are able to receive all frames sent under PCF control. A station that is able to respond to CF-Polls is referred to as being CF-Pollable, and may request to be polled by an active PC. CF-Pollable stations and the PC do not use RTS/CTS in the CFP. If the addressed recipient of a CF transmission is not CF-Pollable, that station acknowledges the transmission using the DCF acknowledgment rules, and the PC retains control of the medium.

13.3 Performance Comparison

This section provides a comparison of DCF and PCF in terms of access delay and throughput. Through the comparisons and simulation results, characteristics of DCF and PCF will be shown out more clearly.

In the first considered scenarios, simulation results in [4] compare average WLAN delays of PCF and DCF stations. Since they have no contention when accessing the medium and need less number of retransmissions, the delays experienced by the packets received by PCF stations are significantly lower than delays of DCFs packets as in Fig. 13.4. In addition, PCF stations also observe less variation in delay values of the received packets, which can be the main quality requirement for some application types. Fluctuation in DCF's delay due to the various backoff time after a fail transmission. However, all stations in the simulation are active and the inactive STAs are not mentioned.

In another approach, the theoretical analysis in [5] illustrates that in a practical environment PCF meets a problem about polling overhead. Especially, it has lots of inactive STAs in the range of PC. Because PC is a central coordinator that schedules channel access for all other pollable stations, and maintains a list of pollable nodes. At the beginning of CFP, it polls all stations in Round-Robin fashion. STAs receiving poll respond back, either by transmitting data or null data frame. Thus, lots of inactive stations cause significant overhead due to lots of null frames transmitted. Overhead time can be computed as follows:

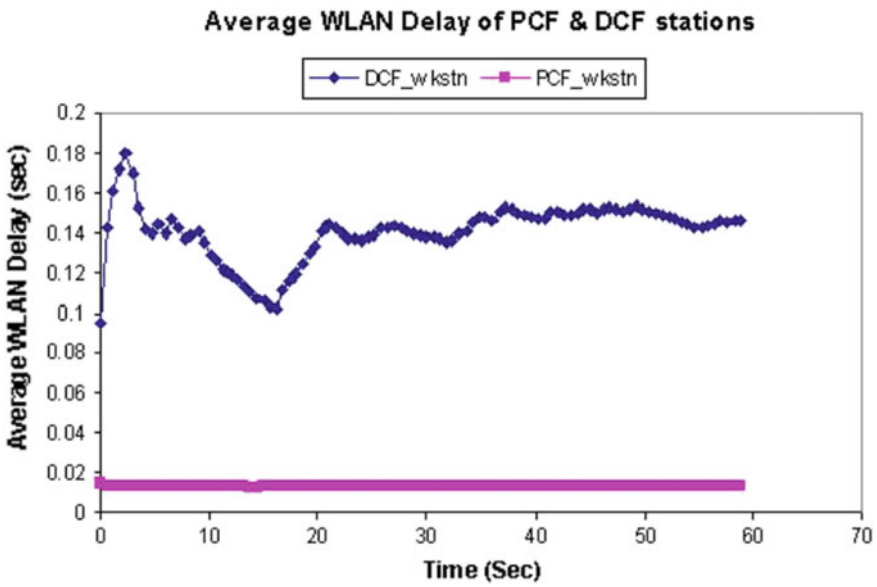


Fig. 13.4 Average WLAN delay for a PCF against DCF stations

$$T_{\text{PollFail}} = T_{\text{Poll}} + \text{SIFS} + T_{\text{Null}} + \text{SIFS} \tag{1.1}$$

Polling overhead also is affected by packet size: the lager packet size, the lower polling overhead. Because with large packet size active STAs seize channel in longer time and then the ratio of T_{PollFail} is reduced. Each medium access control method has an advantage over the other in the relative contrast scenarios. PCF takes higher performance in a high density of transmissions than DCF, but the environment has lots of inactive STAs, PCF meets pooling overhead problems that cause wasting the medium access time as well as processing and energy resources of PC, STAs. Whereas, DCF takes less resources, has better delay in low density of transmission regardless of the number of inactive nodes, but due to the contention, performance of DCF will drop down when the number of active node increases.

13.4 Conclusions

This chapter detailed the characteristics of DCF and PCF—two important channel access mechanisms in IEEE 802.11. Overview about 802.11 MAC is introduced first, then PCF and DCF mechanism is described in detail in Sect. 13.2. Also, this chapter presented comparisons between DCF and PCF in two scenarios. Simulation result as well as theoretical analysis result highlights the advantages and problems of access mechanisms in each case. Through this comparison, new studies about

optimized coexistence of DCF and PCF in IEEE 802.11 can be investigated. PCF need awareness about the inactive nodes for better performance, and the adaptive contention period can take advantages of DCF mechanism.

References

1. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std., 2012
2. Youssef M, Vasan A, Miller R (2002) Specification and analysis of the DCF and PCF protocols in the 802.11 standard using systems of communicating machines. In: 10th IEEE international conference on network protocols, pp 132–141, Nov 2002
3. Dong X, Ergen M, Varaiya P, Puri A (2003) Improving the aggregate throughput of access points in IEEE 802.11 wireless LANs. In: 28th annual IEEE international conference on local computer networks, pp 682–690, Oct 2003
4. Suzuki T, Tasaka S (2001) Performance evaluation of priority-based multimedia transmission with the PCF in an IEEE 802.11 standard wireless LAN. In: 12th IEEE international symposium on personal, indoor and mobile radio communications, vol 2, pp 70–77
5. Goliya A (2003) Dynamic adaption of DCF and PCF mode of IEEE 802.11 WLAN. Ph.D. dissertation, School of Information Technology, Indian Institute of Technology, Bombay

Chapter 14
An Overview of Ultra-Wideband Technology and Its Applications

14.1 Introduction

Ultra-Wideband (UWB) is a technology anticipated to dominate the home networking market and eventually provide carriers with an inexpensive LAN alterative. It offers very high data rates, low power, less expensive cost [1, 2]. UWB provides 100 times the data speeds of Bluetooth solution and allows transmission of large amounts of data i.e. video files between TVs or PCs as well as enabling high quality video applications for portable devices [3–5].

Bluetooth, which named after the tenth century Danish King Harold Bluetooth, is a hot topic among wireless developer. It was designed to allow low bandwidth wireless connections to become to use simply and integrate seamlessly within short range (10 m). Bluetooth wireless technology is the simple choice for wireless, short-range, convenient communications between devices. It is a globally available standard that wirelessly connects mobile phones, portable computers, cars, stereo headsets, MP3 players, and more.

There are more than 50 companies making UWB chips worldwide, including Intel Corp. UWB is a very significant technology but it faces serious regulatory hurdles as well. It is hard for UWB to move forward. The U.S. is the only country to approve spectrum for use by UWB radios. Ultimately, the success of UWB will depend on its low cost. With higher bandwidth, UWB will be adopted in enterprise wireless Personal Area Network (PAN). With appropriate technical standards, UWB devices can operate using spectrum occupied by existing radio services without causing interference, thereby permitting scarce spectrum resources to be used more efficiently. UWB will either become a new age communication or the end of an old technology, and probably both will stay. This chapter presents the overview of UWB technology and its potential application, and the UWB regulation standard worldwide. The UWB short impulse and advantages/disadvantages UWB are also presented.

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_14

14.2 History and Background

Ultra-wideband communications is fundamentally different from all other communication techniques because it employs extremely narrow RF pulses to communicate between transmitters and receivers. Utilizing short-duration pulses as the building blocks for communications directly generates a very wide bandwidth and offers several advantages, such as large throughput, covertness, robustness to jamming, and coexistence with current radio services.

Ultra-wideband communications is not a new technology; in fact, it was first employed by Guglielmo Marconi in 1901 to transmit Morse code sequences across the Atlantic Ocean using spark gap radio transmitters. However, the benefit of a large bandwidth and the capability of implementing multiuser systems provided by electromagnetic pulses were never considered at that time.

Approximately fifty years after Marconi, modern pulse-based transmission gained momentum in military applications in the form of impulse radars. Several pioneers of modern UWB communications in the United States from the late 1960s was establishes such as Henning Harmuth of Catholic University of America and Gerald Ross and K. W. Robins of Sperry Rand Corporation [6]. From the 1960s to the 1990s, this technology was restricted to military and Department of Defense (DoD) applications under classified programs such as highly secure communications. However, the recent advancement in micro processing and fast switching in semiconductor technology has made UWB ready for commercial applications. Therefore, it is more appropriate to consider UWB as a new name for a long-existing technology.

As interest in the commercialization of UWB has increased over the past several years, developers of UWB systems began pressuring the FCC to approve UWB for commercial use. In February 2002, the FCC approved the First Report and Order (R&O) for commercial use of UWB technology under strict power emission limits for various devices. Figure 14.1 summarizes the development timeline of UWB.

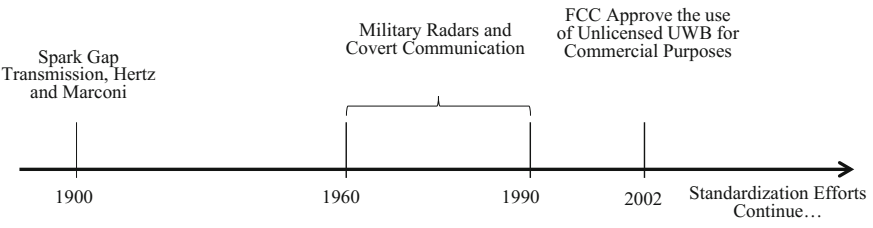


Fig. 14.1 A brief history of UWB developments

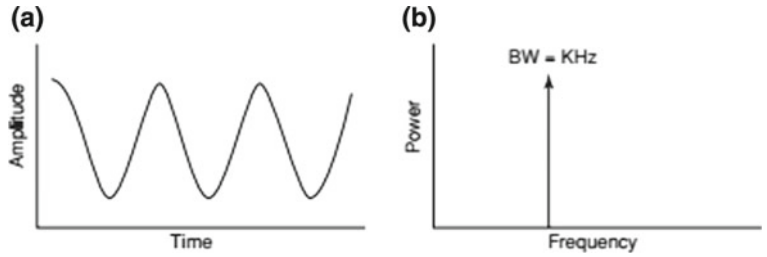


Fig. 14.2 A narrowband signal in **a** the time domain and **b** the frequency domain

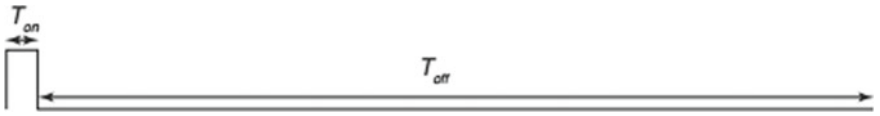


Fig. 14.3 A low-duty-cycle pulse. T_{on} represents the time that the pulse exists and T_{off} represents the time that the pulse is absent

14.3 UWB Concepts

Traditional narrowband communications systems modulate continuous waveform (CW) RF signals with a specific carrier frequency to transmit and receive information. A continuous waveform has a well-defined signal energy in a narrow frequency band that makes it very vulnerable to detection and interception. Figure 14.2 represents a narrowband signal in the time and frequency domains.

As mentioned in Sect. 12.2, UWB systems use carrierless, short-duration (picosecond to nanosecond) pulses with a very low duty cycle (less than 0.5%) for transmission and reception of the information. A simple definition for duty cycle is the ratio of the time that a pulse is present to the total transmission time. Figure 14.3 and Eq. 14.1 represent the definition of duty cycle.

$$\text{Duty cycle} = \frac{T_{\text{on}}}{T_{\text{on}} + T_{\text{off}}} \tag{14.1}$$

Low duty cycle offers a very low average transmission power in UWB communications systems. The average transmission power of a UWB system is on the order of microwatts, which is a thousand times less than the transmission power of a cell phone! However, the peak or instantaneous power of individual UWB pulses can be relatively large, but because they are transmitted for only a very short time ($T_{\text{on}} < 1$ ns), the average power becomes considerably lower. Consequently, UWB devices require low transmit power due to this control over the duty cycle, which directly translates to longer battery life for handheld equipment. Since frequency is inversely related to time, the short-duration UWB pulses spread their energy across a wide range of frequencies—from near DC to several gigahertz (GHz)—with very

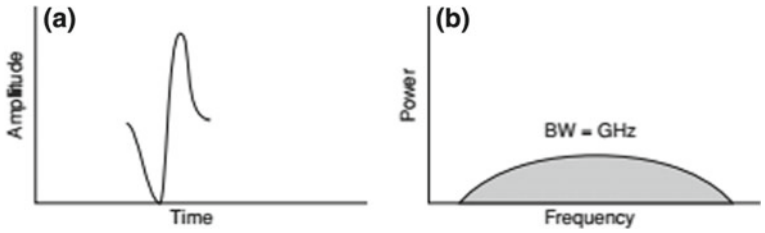


Fig. 14.4 A UWB pulse in **a** the time domain and **b** the frequency domain. Compare the bandwidth and power spectral density with those of the narrowband signal in Fig. 14.2

low power spectral density (PSD). Figure 14.4 illustrates UWB pulses in time and frequency domains.

UWB technology has the following significant characteristics [8].

14.3.1 High Data Rate

UWB can handle more bandwidth-intensive applications like streaming video, than either 802.11 or Bluetooth because it can send data at much faster rates. UWB technology has a data rate of roughly 100 Mbps, with speeds up to 500 Mbps, This compares with maximum speeds of 11 Mbps for 802.11b (often referred to as Wi-Fi) which is the technology currently used in most wireless LANs; and 54 Mbps for 802.11a, which is Wi-Fi at 5 MHz. Bluetooth has a data rate of about 1 Mbps (Fig. 14.5).

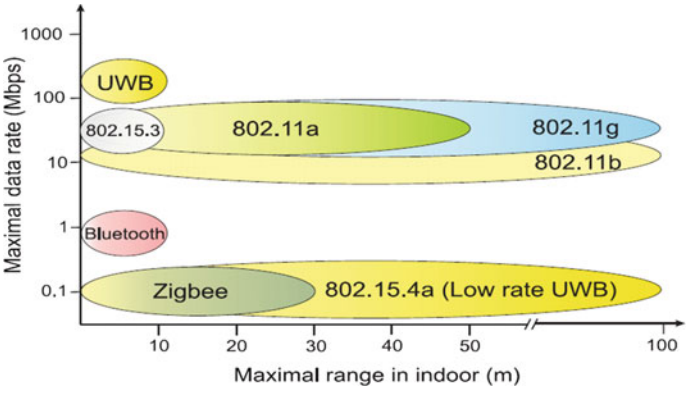


Fig. 14.5 Maximum range and data rate of different wireless technologies

14.3.2 Low Power Consumption

UWB transmits short impulses constantly instead of transmitting modulated waves continuously like most narrowband systems do. UWB chipsets do not require Radio Frequency (RF) to Intermediate Frequency (IF) conversion, local oscillators, mixers, and other filters. Due to low power consumption, battery-powered devices like cameras and cell phones can use in UWB.

14.3.3 Interference Immunity

Due to low power and high frequency transmission, USB's aggregate interference is "undetected" by narrowband receivers. Its power spectral density is at or below narrowband thermal noise floor. This gives rise to the potential that UWB systems can coexist with narrowband radio systems operating in the same spectrum without causing undue interference.

14.3.4 High Security

Since UWB systems operate below the noise floor, they are inherently covert and extremely difficult for unintended users to detect.

14.3.5 Reasonable Range

IEEE 802.15.3a Study Group defined 10 m as the minimum range at speed 100 Mbps However, UWB can go further. The Philips Company has used its Digital Light Processor (DLP) technology in UWB device so it can operate beyond 45 ft at 50 Mbps for four DVD screens.

14.3.6 Large Channel Capacity

The capacity of a channel can be express as the amount of data bits transmission/second. Since, UWB signals have several gigahertz of bandwidth available that can produce very high data rate even in gigabits/second. The high data rate capability of UWB can be best understood by examining the Shannon's famous capacity equation:

$$C = B * \log_2\left(1 + \frac{S}{N}\right), \tag{14.2}$$

where C is the channel capacity in bits/second, B is the channel bandwidth in Hz, S is the signal power and N is the noise power. This equation tells us that the capacity of a channel grows linearly with the bandwidth W, but only logarithmically with the signal power S. Since the UWB channel has an abundance of bandwidth, it can trade some of the bandwidth against reduced signal power and interference from other sources. Thus, from Shannon's equation we can see that UWB systems have a great potential for high capacity wireless communications.

14.3.7 Low Complexity, Low Cost

The most attractive of UWB's advantages are of low system complexity and cost. Traditional carrier based technologies modulate and demodulate complex analog carrier waveforms. In UWB, Due to the absence of Carrier, the transceiver structure may be very simple. The techniques for generating UWB signals have existed for more than three Decades. Recent advances in silicon process and switching speeds make UWB system as low-cost. Also home UWB wireless devices do not need transmitting power amplifier. This is a great advantage over narrowband architectures that require amplifiers with significant power back off to support high-order modulation waveforms for high data rates.

14.3.8 Resistance to Jamming

The UWB spectrum covers a huge range of frequencies. That's why, UWB signals are relatively resistant to jamming, because it is not possible to jam every frequency in the UWB spectrum at a time. Therefore, there are a lot of frequency range available even in case of some frequencies are jammed.

14.3.9 Scalability

UWB systems are very flexible because their common architecture is software redefinable so that it can dynamically trade-off high-data throughput for range.

14.4 UWB Technologies

In the near future this technology may see increased use for high-speed short range wireless communications, ranging and ad hoc networking. There are two competing technologies for the UWB wireless communications, namely: Impulse Radio (IR) and Multi-band OFDM (MB-OFDM). IR technique is based on the transmission of very short pulses with relatively low energy. The MB-OFDM approach divides the UWB frequency spectrum to multiple non-overlapping bands and for each band transmission is OFDM. Several proposals based on these two technologies have been submitted to the IEEE 802.15.3a. Both technologies are valid and credible.

14.4.1 Impulse Radio

In impulse radio UWB pulses of very short duration (typically in the order of sub-nanosecond) are transmitted. Because of very narrow pulses the spectrum of the signal reaches several GHz of bandwidth. The impulse radio UWB is a carrier-less transmission. This technology has a low transmit power and because of narrowness of the transmitted pulses has a fine time resolution. The implementation of this technique is very simple as no mixer is required which means low cost transmitters and receivers. Direct Sequence Ultra-Wideband (DS-UWB) and Time Hopping Ultra-Wideband (TH-UWB) are two variants of the IR technique. These IR techniques DS-UWB and TH-UWB are different multiple access techniques that spread signals over a very wide bandwidth. Because of spreading signals over a very large bandwidth, the IR technique can combat interference from other users or sources. It should be mentioned that Direct Sequence Spread Spectrum (DSSS) and Time Hopping Spread Spectrum (THSS) may be considered similar to DS-UW Band TH-UWB, respectively. There are, however, differences between the spread spectrum and IR-UWB systems. Both systems take advantage of the expanded bandwidth, while different methods are used to obtain such large bandwidth. In the conventional spread-spectrum techniques, the signals are continuous-wave sinusoids that are modulated with a fixed carrier frequency, while in the IR-UWB (i.e., DS-UWB and TH-UWB), signals are basically baseband and the narrow UWB pulses are directly generated having an extremely wide bandwidth. Another difference is the bandwidth. For the UWB signals the bandwidth has to be higher than 500 MHz, while for the spread spectrum techniques bandwidths are much smaller (usually in the order of several MHz).

14.4.2 Multiband OFDM

Multi-band Orthogonal Frequency Division Multiplexing (MB-OFDM) is another UWB technology which uses the OFDM method. Multi-Band OFDM combines the

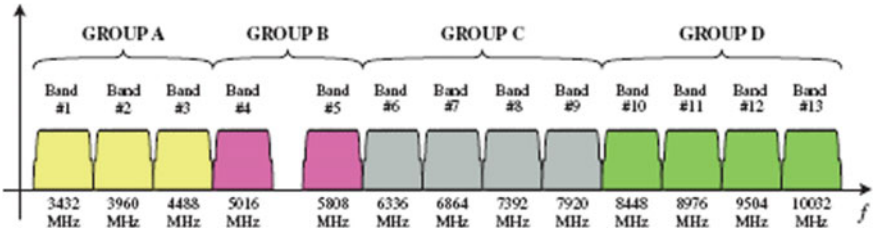


Fig. 14.6 Proposed MB-OFDM frequency band plan

OFDM technique with the multi-band approach. The spectrum is divided into several sub-bands with a −10 dB bandwidth of at least 500 MHz (Fig. 14.6).

The information is then interleaved across sub-bands and then transmitted through multi-carrier (OFDM) technique. One of the proposals for the physical layer standard of future high speed Wireless Personal Area Networks (WPANs) uses MB-OFDM technique. In this MB-OFDM WPANs proposal, the spectrum between 3.1 and 10.6 GHz is divided into 14 bands with 528 MHz bandwidth that may be added or dropped depending upon the interference from, or to, other systems. In Fig. 14.6 a possible band plan is presented, where only 13 bands are used to avoid interference between UWB and the existing IEEE 802.11a signals. The three lower bands are used for standard operation, which is mandatory, and the rest of the bands are allocated for optional use or future expansions. MB-OFDM technology promises to deliver data rates of about 110 Mbps at a distance of 10 m. For the UWB wireless sensor applications data rates are low, but the (hoping) coverage might be much larger than 10 m. MB-OFDM may require higher power levels when compared to the IR technology. MB-OFDM technique is robust to multipath which is present in the wireless channels.

The advantages of the MB-OFDM technique are as follows:

- Capturing multipath energy with a single RF chain
- Insensitivity to group delay variations
- Ability to deal with narrowband interference at receivers
- Simplified synthesizer architectures relaxing the band switching timing requirements.

The disadvantages are as follows:

- Transmitter is more complex because of IFFT
- High peak-to-average power ratios
- OFDM synchronization problems.

Table 14.1 Specifications comparison of MB-OFDM and IR DS-UWB techniques for WPAN

Specifications	MB_OFDM	IR(DS-UWB)
Number of sub-bands	3 mandatory, up to 14	2 (3.1–4.85 GHz and 6.2–9.7 GHz)
Sub-band bandwidth	528 MHz	1.75 GHz (lower band) 3.5 GHz (higher band)
Number of sub-carriers	122	No sub-carriers (baseband signals)
Spreading factor	1, 2	1–24
Data rates (Mbps)	55, 80, 110, 160, 200, 320, 480	28, 55, 110, 220, 500, 660, 1000, 1320 (lower band)
Modulation	QPSK	BPM, MBOK
Multiple access	Based on time-frequency codes	Based on PN codes

14.4.3 Comparison of UWB Technologies

The comparison of IR DS-UWB and MB-OFDM UWB techniques in terms of interference from, or to, other systems, robustness to multipath, performance, system's complexity and achievable range-data rate performance for the WPAN applications is provided as in Table 14.1.

14.5 Technologies and Standards

14.5.1 Bluetooth

Bluetooth [7] is a globally available standard that wirelessly connects mobile phones, portable computers, cars, stereo headsets, MP3 players, and more. It is an ad hoc technology that requires no fixed infrastructure and is simple to install and set up. Since the first release of the Bluetooth specification in 1999, over 4000 companies have become members in the Bluetooth Special Interest Group (SIG). Meanwhile, the number of Bluetooth products on the market is multiplying rapidly. A simple example of a Bluetooth application is updating the phone directory of your mobile phone. You would have to either manually enter the names and phone numbers of all your contacts or use a cable or IR link between your phone and your PC and start an application to synchronize the contact information. With Bluetooth, this could all happen automatically and without any user involvement as soon as the phone comes within range of the PC.

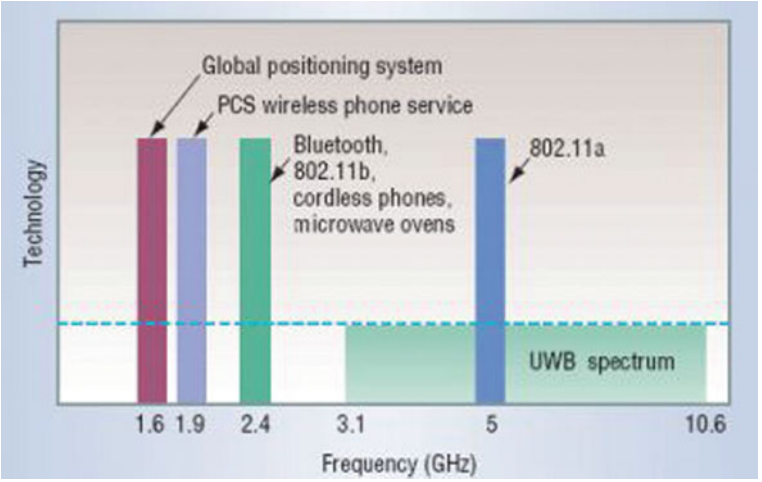


Fig. 14.7 Wireless technology frequencies

14.5.2 UWB

UWB is designed to replace cables with short-range, wireless connections, but it offers much higher bandwidth needed to support huge amounts of data streams at very low power levels [9]. Examples include media players, monitors, cameras, and cell phones. Because UWB can communicate both relative distance and position, it can be used for tracking equipment, containers or other objects. UWB chipsets are built in complementary metal oxide semiconductor, so they rival inexpensive Bluetooth price when produced in volume. A recent technology demonstration showed a UWB device transmitted at a data rate of 110 Mbit/s at a range of up to 10 m.

As shown in Fig. 14.7, unlike conventional radio systems, which operate within a relatively narrow bandwidth, ultra-wideband operates across a wide range of frequency spectrum by transmitting a series of extremely narrow (10–1000 ps) and low power pulses. The possible use of UWB technology in communications ranges from WLAN-like office or home networking and Internet access. By using 80% less power than 802.11a, UWB chipsets can work with smaller device such as PDAs and mobile phones without unduly burdening their batteries.

The primary advantages of UWB are high data rates, low cost, and low power. Because UWB is spectrum hopping, and only for a tiny fraction of a second, UWB causes less interference than narrowband radio designs, nearby neighbors will not interfere with other UWB networks. An additional UWB feature, precise ranging, or distance measurement is used for location identification, i.e. tracking persons. UWB uses very little power with long battery life. For the security issue, it is extremely hard to eavesdrop. It is like trying to track someone in a very busy street who continually changes different colors of clothes while running at extreme fast speed. Despite the many benefits of UWB, it is currently embroiled in specification and standardization

Table 14.2 Comparison of UWB and Bluetooth

Technology	UWB	Bluetooth
Spectrum (GHz)	3.1–10.6	2.4
Typical range (m)	10–30	10
Technology	OFDM or DS-UWB	Adaptive frequency-hopping spread spectrum
Max data rate	1 Gbit/s	1 Mbit/s
Typical application	Wireless synchronization and data transmission	Low-bandwidth wireless interconnect for synchronizing data
Availability	After 2007	Now

wrangling within the standard issuing bodies of the world and the USA, namely IEEE, ITU and the FCC (Federal Communications Commission).

The drawback is such speeds only work over short distances. Communication speed is a function of bandwidth, power, and distance. The crossover point for UWB versus 802.11a wireless is 10 m—less than 10 m, and UWB has higher bandwidth, but over 10 m, 802.11a wins. Because of UWB's distance limitations, it will primarily be used for high-bandwidth local networks where the receiver can be plugged in, and not for cellular. Another drawback is that UWB standards battle remains unresolved. It is likely that UWB and Bluetooth could both be integrated into end-devices to serve different application spaces. Table 14.2 compares the both technologies in spectrum, range, data rate, and user applications aspects.

Electromagnetic waves with instantaneous bandwidth greater than 25% of the center operating frequency or an absolute bandwidth of 1.5 GHz or more are referred to be ultra-wideband (UWB) signals. UWB radio systems with bandwidths more than 1.5 GHz but an instantaneous bandwidth less than 25% of the center operating frequency can be designed using traditional RF components (antennas, frequency synthesisers, amplifiers and filters) which are reasonably straightforward to produce. On the other hand, UWB systems designed at lower frequencies (typically less than 3 GHz) with an instantaneous bandwidth greater than 25% of the center operating frequency require a more novel approach.

14.5.3 UWB Standards

14.5.3.1 IEEE 802.15.3

IEEE 802.15.3 is the IEEE standard for high data rate (20 Mbit/s or greater) Wireless Personal Area Networks (WPAN) to provide Quality of Service (QoS) for real time distribution of multimedia content. IEEE 802.15.3 is accomplished by the IEEE P802.15.3 High Rate (HR) Task Group (TG3). The task group is charged with defining a universal standard of ultra-wideband radios capable of high data rate

over a distance of 10 m using the 3.1−10.6 GHz band (see Fig. 14.1) for TVs, cell phones, PCs, and so forth. Besides a high data rate, the new standard will provide for low power, low cost solutions addressing the needs of portable consumer digital imaging and multimedia applications. In addition, ad hoc peer-to-peer networking, security issues are considered. When combined with the 802.15.3 PAN standard, UWB will provide a very compelling wireless multimedia network for the home.

The IEEE 802.15.3 standard enables wireless multimedia applications for portable consumer electronic devices within home coverage. The standard supports wireless connectivity for gaming, printers, cordless phones and other consumer devices. It can be used to develop wireless multimedia applications including wireless surround sound speakers, portable video displays, digital video cameras. It addresses the need for mobility, quality of service (QoS) and fast connectivity for the broad range of consumer electronic devices.

14.5.3.2 WiMedia UWB

The WiMedia Alliance is a nonprofit open industry association that promotes and enables the standardization and multi-vendor interoperability of ultra-wideband worldwide. The new WiMedia Alliance represents a combination of WiMedia with the Multiband OFDM Alliance SIG (MBOA-SIG). Both are two leading organizations. They will publish and manage the industry UWB specifications for rapid adoption by for mobile, consumer electronics and PC applications. The MBOA-SIG Promoter companies include Alereon, HP, Intel, Kodak, Microsoft, Nokia, Philips, Samsung Electronics, Sony, etc. MBOA member companies are actively engaged with IEEE standards process.

The MOBA will announce its specifications for a physical layer (“PHY”) and Media Access Control layer (“MAC”) to enhance personal electronic devices mobility. The MBOA MAC and PHY specifications will serve as the common radio platform for industry standards. The MBOA MAC and PHY specifications, as published in ECMA-368, are intentionally designed to adapt to various requirements set by global regulatory bodies. The Multiband OFDM Alliance (MBOA) has devised its own media access control (MAC) layer, in effect rejecting the MAC mandated by the IEEE for the upcoming 802.15.3a standard. Enhanced support for mobility, mesh networking and management of Piconets will be the key to the new MAC. Other application-friendly features in MBOA include the reduced level of complexity per node, long battery life, support of multiple power management modes and higher spatial capacity.

14.5.3.3 End User Applications

WB has many other applications, for instance, medical imaging, automobile collision-avoidance systems, firefighters and police looking through walls, as well as finding and tracking assets and people. This is a technology which, in at least

some applications, could be saving lives. For MBOA UWB, anticipated early applications include the exchange of media content over high-data consumer electronics devices including MP3 players, personal media players (PMPs), set-top-boxes, digital cameras, hard-drives, printers/scanners, home-theater equipment, mobile phones, personal computers and video gaming platforms.

14.5.4 Marketplace and Vendor Strategies

The two sides in UWB standards battle are more polarized in wireless personal area networking market. It is sort of hard on both end users and vendors. It's obvious how end users suffer. They have to gamble on a standard proposal that might lose. For enterprise users, the risk may be unacceptable, leaving them, not with two (or three) options, but with none. If one of the proposals wins, the companies that were involved in the losing proposal certainly take a serious hit. All their development time and investment is gone, and they have to design and build new chips. This could take two years or so. Even the winners have extra, competitive pressure from the desire to set a standard means they probably haven't been able to make as much money as they otherwise might off of a slow but steady start in a more cohesive marketplace. Marketing and promotion expenses are high for both groups.

One way of resolving a conflict that doesn't seem to be getting any better through the normal standards process is to let the parties fight it out in the marketplace. That is to say, let them ship products, and see which ones eventually win. The risk of picking the wrong standard creates a real reason to adopt a "wait and see" attitude towards new standards. The basic advice for standards battles is to stay clear. Waiting until the dust settles a bit and you can tell what the standard is before adopting one.

14.6 UWB Applications

Although it is claimed that many exotic applications would benefit from UWB technology, the literature search revealed that there are two main potential UWB application areas: communications and radar/sensor. For both areas, the basic UWB system components include transmitter sources, modulators, RF pulse generators, detection receivers and wideband antennas. There has been a significant amount of research into the development of UWB components over many years. It is suggested that the antenna design remains to be a significant challenge. The options being considered include loaded dipoles, TEM horns, biconicals and ridged horns, spiral and large current antennas, each with a variety of advantages and disadvantages.

The high-data-rate capability of UWB systems for short distances has numerous applications for home networking and multimedia-rich communications in the form of WPAN applications. UWB systems could replace cables connecting camcorders and VCRs, as well as other consumer electronics applications, such as laptops, DVDs,

Table 14.3 UWB capabilities compared to other IEEE standards

IEEE Standard							
	WLAN			Bluetooth	WPAN	UWB	ZigBee
	802.11a	802.11b	802.11 g	802.15.1	802.15.3	802.15.3a	802.15.4
Frequency (GHz)	5	2.4	2.4	2.4	2.4	3.1–10.6	2.4
Max data rate	54 Mbps	11 Mbps	54 Mbps	1 Mbps	55 Mbps	>100 Mbps	250 Kbps
Max range (m)	100	100	10	10	10	10	5

digital cameras, and portable HDTV monitors. No other available wireless technologies—such as Bluetooth or 802.11a/b—are capable of transferring streaming video. Table 14.3 compares UWB technology and other currently available data communications standards.

In fact, many of the current communications and radar/sensor devices are band limited due largely to the bandwidth limitations imposed by the antennas which act as band pass filters in UWB transmissions. More recent system proposals do not rely on band limited transmissions which, in turn, brings about the requirements for modification. The following sub-sections summaries the likely communications and radar/sensor applications.

14.6.1 Communications

It is argued that the types of potential communications devices to be deployed will largely be dictated by the emission limits enforced by regulatory authorities. The maximum operating distance and the transmission rate will be the key parameters for the performance assessment of the UWB communications systems. Operational characteristics of some of the devices are outlined below.

Handheld transceiver designed for full duplex voice and data transmissions up to 128 kbps, operating at 1.5 GHz with an instantaneous bandwidth of 400 MHz's. The peak output power is measured to be 2 W and the LOS range is up to 2 km. With small gain antennas, the range extends up to 32 km. Ground wave communications system designed for non-LOS digital voice and data transmission up to 128 kbps, operating in the 30–50 MHz band over a range of 16 km with a peak power of approximately 35 W. Asymmetric, bi-directional video/command and control UWB transceivers designed to operate in the range 1.3–1.7 GHz with transmission rates up to 25 Mbps using 4 W peak output power. Handheld transceivers designed for multichannel, full duplex, 32 kbps digital voice transmissions over a range of 100 m

in the band 1.2–1.8 GHz on board a navy craft. Indoor short range communications device operating in the range 2.5–5 GHz over a range of 50 m providing data rates up to 60 Mbps.

14.6.2 Radars/Sensors

The primary use of UWB radars is to provide target detection while the UWB sensors are used to obtain information concerning the target [8]. The number of applications is extensive. These include ground penetration, position location, and wall penetration, collision warning for avoidance, fluid level detection, intruder detection and vehicle radar. New applications include distance and air-bag proximity measurements, road and runway inspection, heart monitoring, RF identification and camera auto focus. System characteristics of some of the radars/sensors are summarized below.

Vehicular Electronic Tagging and Alert System designed to relay the picture of the driver together with information on the driver and the vehicle to a roadside sensor in a police vehicle. The system operates in 1.4–1.65 GHz region with a peak power of 0.25 W over a range of 300 m. Geolocation system designed to provide three dimensional location information, operating in 1.3–1.7 GHz region by utilizing 2.5 ns, 4 W peak power UWB pulses. LOS range is 2 km with omnidirectional antennas. Indoor range is up to 100 m.

Precision altimeter and collision avoidance sensor designed to operate in the 5.4–5.9 GHz range with peak output power of 0.2 W. Backup sensor designed to detect objects behind large construction and mining vehicles, operating with 0.25 W peak power in 5.4–5.9 GHz region over a range up to 100 m. Electronic license plate designed to provide both automobile collision avoidance and RF tagging for vehicle to roadside communications. Collision avoidance functions are provided using 0.2 W peak power in 5.4–5.9 GHz region over a 30 m range while the tagging functions are supported with a 0.3 W peak power over a range 200 m. Military radar designed for very short range applications (less than 2 m) with an average power of 85 nW operating at 10 GHz with a 2.5 GHz bandwidth.

14.7 Conclusions

Federal Communications Commission gave its approval to sell UWB wireless products in the U.S. Although the lack of an adopted standard will slow growth for a while, UWB will be used in 150 million devices by 2008. Since UWB is best used for short-distance and high-bandwidth applications, most of the development in UWB is targeted at HDTV and DLP video projection. UWB is not seen or designed as a replacement of traditional Wi-Fi. WiMax and MobileFi are seen as that replacement but that is another story.

References

1. Tran M-P, Minh Thu PT, Lee H-M, Kim D-S (2015) Effective spectrum handoff for cognitive UWB industrial networks. In: ETFA WIP, Luxembourg, 8–11 Sept 2015
2. Siwiak K (2001) Ultra-wide band radio: introducing a new technology. In: IEEE VTS 53rd vehicular technology conference, vol 2, pp 1088–1093
3. Fontana R (2004) Recent system applications of short-pulse ultra-wideband (UWB) technology. IEEE Trans Microw Theory Tech 52(9):2087–2104
4. Colson S, Hoff H (2005) Ultra-wideband technology for defence applications. In: IEEE international conference on ultra-wideband, pp 615–620, Sept 2005
5. Immoreev I, Fedotov P (2002) Ultra wideband radar systems: advantages and disadvantages. In: IEEE conference on ultra wideband systems and technologies, pp 201–205, May 2002
6. Lahaie IJ (ed) (1992) Ultrawideband radar: SPIE proceedings, vol 1631, Jan 1992
7. Bisdikian C (2001) An overview of the bluetooth wireless technology. IEEE Commun Mag 39(12):86–94
8. Patni ML (2004) Ultra-wideband: the next generation personal area network technology. Computer Systems Limited, Feb 2004
9. Lynch A, Genello B, Wicks C (2007) UWB perimeter surveillance. IEEE Aerosp Electron Syst Mag 22(1):8–10

Chapter 15
Ultra-Wideband Technology for Military Applications

15.1 Introduction

Ultra-WideBand (UWB) communication systems are usually classified as any communication system whose instantaneous bandwidth is many times greater than the minimum bandwidth required to deliver information. UWB communication is fundamentally different from all other communication techniques because it employs extremely narrow Radio Frequency (RF) pulses to communicate between transmitters and receivers. Utilizing short-duration pulses as the building blocks for communications directly generates a very wide bandwidth and offers several advantages, such as large throughput, covertness, robustness to jamming, and coexistence with current radio services. UWB technology offers a promising solution to the RF spectrum drought by allowing new services to coexist with current radio systems with minimal or no interference. This coexistence brings the advantage of avoiding the expensive spectrum licensing fees that providers of all other radio services must pay.

The first wireless transmission via UWB emissions was sent by Marconi from the Isle of Wight to Cornwall on the British Mainland in 1901 using Marconi Spark Gap Emitter [1]. The UWB signal was created by the random conductance of a spark. Then in the late 1970s and early 80s, Fullerton demonstrated the practicality of modern low power impulse radio techniques using time-coded time-modulated ultra-wideband approach.

The Federal Communications Commission (FCC) is the RF controlling body in the United States. It controls spectrum division and licensing. It had earlier committed UWB to experimental work only, commercial use was not allowed. In 2002, the FCC changed the rules to allow UWB system operation in a broad range of frequencies. In 2003, the first FCC-certified commercial system was installed, and in April 2003, the first FCC-compliant commercial UWB chipsets were announced by Time Domain Corporation.

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_15

Signal is defined to be an ultra-wideband if the fractional bandwidth B_f is greater than 0.25 [2]. The fractional bandwidth can be determined using the following formula:

$$B_f = 2\frac{f_H - f_L}{f_H + f_L}, \tag{15.1}$$

where

B_f is fractional bandwidth,
f_L is lower, and
f_H is higher 3 dB point in a spectrum, respectively.

15.2 Technical Overview of Ultra-Wideband Systems

Ultra-Wideband is a technology for the transmission of data using techniques which cause a spreading of the radio energy over a very wide frequency band, with a very low power spectral density. The low power spectral density limits the interference potential with conventional radio systems, and the high bandwidth can allow very high data throughput for communications devices, or high precision for location and imaging devices.

UWB technique is a radio transmission technology which occupies a relatively wide bandwidth, which exceeds 500 MHz as a minimum or it has at least 20% of the center frequency. This technology has gained attention because of its potential as a novel approach for short-range and wide bandwidth wireless communication. In contrast to the traditional narrowband communication systems, UWB systems transmit information by generating radio energy at specific time instants in the form of very short pulses. Thus, these systems occupy a relatively large bandwidth and enable us to use time modulation. In addition, UWB systems can provide a high data rate which can reach up to hundreds of Mbps, and which make them useful for secure communication in military applications. Moreover, the UWB systems demand a relatively low transmitting power in comparison with the traditional narrowband communications systems; hence, their use can prolong the battery life. In addition, use of short pulses helps reduce multipath channel fading since the reflected signals do not overlap with the original ones.

Most of the time, we refer to all types of UWB systems with a single name, but in reality, there are two very different technologies being developed:

- **Carrier free direct sequence ultra-wideband technology**: This form of ultra-wideband technology transmits a series of impulses. In view of the very short duration of the pulses, the spectrum of the signal occupies a very wide bandwidth.
- **MBOFDM (Multi-Band OFDM ultra-wideband technology)**: This form of ultra-wideband technology uses a wide band or Multi-Band Orthogonal Frequency Division Multiplex (MBOFDM) signal that is effectively a 500 MHz wide OFDM

signal. This 500 MHz signal is then hopped in the frequency to enable it to occupy a sufficiently high bandwidth.

The working principles of UWB technology is first a signal with ultra-wide bandwidth is generated using electrical short, baseband pulses (100 ps–1 ns). Then, the data is transmitted using Pulse Modulation (Amplitude, Position, or Phase Modulation), and then the baseband pulses are directly applied to the antenna and finally, a correlation receiver or rake receiver is used to capture the signal.

15.3 Ultra-Wideband Technology for Military Applications

Recently, ultra-wideband (UWB) technology has attracted much attention both in the industry, military, and academic area due to its low cost, potential to handle high data rate and relatively low power requirement. UWB has a wide range of applications; most recent applications target sensor data collection, precision locating, and tracking applications. There are a wide number of applications that UWB technology can be used for. They vary from data and voice communications through to radar and classification. Although much of the excitement about ultra-wideband has been associated with commercial applications, the technology is equally suited to military applications. The use of a new technology for military applications requires a rigorous, practical, and objective analysis in order to point out the eventual advantages of this technology compared with the already used solutions.

One of the missions of CELAR [3] (Technical Center for Armament Electronics, which belongs to DGA—French Procurement Agency) is to participate to UWB analysis for the French Ministry of Defense (MoD). The concrete military applications are deduced and illustrated by studies currently held at CELAR. The operative missions for which Impulse Radio UWB technology had been implemented could bring an improvement to achieve different missions using a single technology. The inherent properties of IR-UWB are quite interesting, even though some are problematic [4]. There are many potential advantages of UWB in Commercial, Industrial and Educational environments. Moreover, currently, UWB applications are quite interesting for military applications mainly for the following reasons.

- *Higher data rate*: UWB can increase capacity while maintaining low power transmission. The different manufacturers are talking about data rates of 140 Mbps for small range wireless communications, compared with data rates of existing solutions that are five times lower (20 Mbps for WiFi, 2 Mbps for Universal Mobile Telecommunication Systems (UMTS), and 1 Mbps for Bluetooth).
- *Robustness to multipath fading*: The scattering Radar Cross-Section (RCS) of these interfering sources is reduced with respect to the target RCS, because of the small spatial extent of the pulse. Moreover, due to the very large bandwidth of signal, a significant part of the transmitting energy propagates at wavelengths for which rain, mist, or aerosols are only lightly absorbing. If the transmitter and the receiver are not at the same place (usual case for communication applications), as UWB

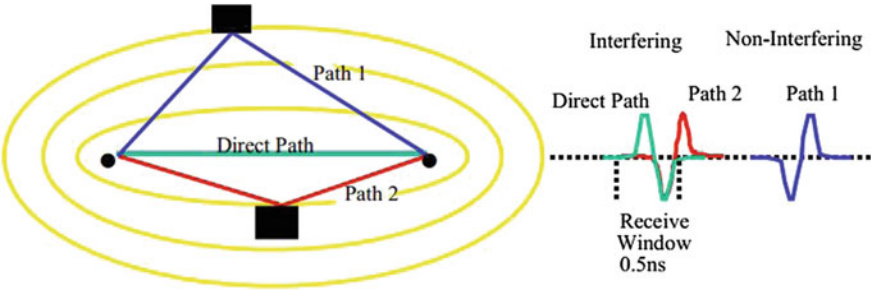


Fig. 15.1 Illustration of the robustness of the UWB signal to multipath fading for mobile communication applications

pulses are very short (about 1 ns, even less), the direct path has come and gone before the reflected path arrives (Fig. 15.1).

- *Low probability of detection (LPD) and Low probability interception (LPI)*: As a very short-duration pulse implies a large band, the power is spread over numerous frequencies instead of being concentrated. The resultant power spectral density is very low. Consequently, the probability of detection and interception is very low. This is a very useful point for military applications. Moreover, due to this low power level, a minimum power exposure for the user may occur.

It is known that the UWB pulse is generated in a very short time period (sub-nano second). This technology is young and still needs to attract more researchers and engineers. Because of its broad applicability, today there are thousands of applications where UWB radars are used. As a result of its relevancy, indifferent areas such as academic institutions, government and private sectors, and defenses and military, they are making promising researches to apply unique features of UWB technology. Research into landmine detection using UWB radar took place in the UK during the Falklands conflict in the early 1980s and in the U.S. in the aftermath of the Gulf War a decade later. Military research programmers in Germany, France, and Russia followed, but until now, in the wake of pioneering research into the feasibility of UWB microwave imaging for early-stage breast cancer detection that the subject has started to gain serious traction on both sides of the Atlantic.

UWB microwave imaging uses the measurement of the transmission of microwave energy through an object to define its dielectric (insulating) properties. Put simply, by subjecting the tissue to signals from across the radio spectrum at a much wider range than traditional X-rays, scientists at the University of Wisconsin–Madison in the U.S. have been able to contrast normal fatty breast tissue with malignant tissue. This not only makes it possible to identify extremely small malignant tumors but also to differentiate between benign and malignant growths.

Military chiefs hope to adapt the UWB microwave technology to detect improvised explosive device (IEDs) in the field. A recent declassified NATO report on research into handheld, standoff vehicle-based, and airborne detection systems outlined some of the current and future engineering challenges affecting IED detection,

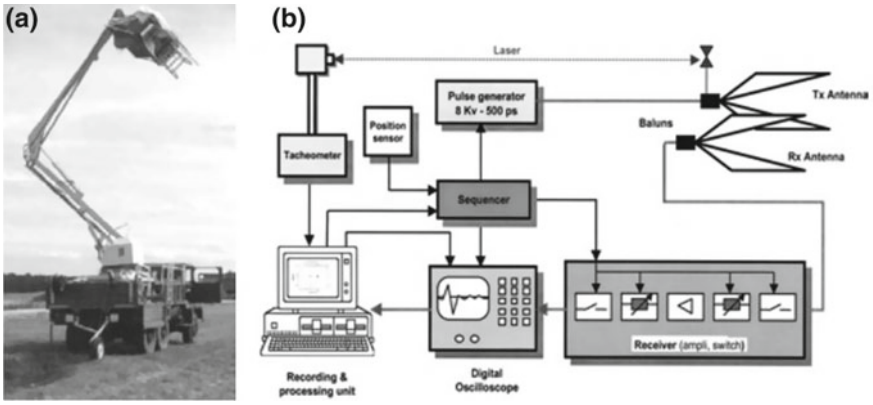


Fig. 15.2 PULSAR radar (**a**) and its architecture (**b**)

recognition, and identification using UWB microwaves. "UWB radar has emerged as the technology of choice for improved landmine detection," stated the report by UK specialist technical consultancy the Electrical Research Association (ERA) Technology. "Recent developments using handheld dual sensor technology combining Electromagnetic Induction (EMI) and Ground-Penetrating Radar (GPR) have enabled improved discrimination against metallic fragments to be demonstrated in live minefields.

The applications and contributions of UWB technology on military areas through radar systems (PULSAR radar) have been studied by [5, 6]. The features of UWB radar are quite increasing because of its ability to detect through obstacles and dense media; which can be useful in military field and rescue operations for detecting buried people under (snow, rock, and mining), improved clutter rejection, improvements of radar range resolution, improved detection of low flying targets, and improved recognition of targets. These all applications are related with the high power UWB.

The low power UWB could also be used for the civilian and military needs. Low power UWB could be used for civilian purpose to provide high data rate domestic networks. The most important feature of low power UWB technology is for military needs using IR-UWB in the intra-squad communication system. This could provide localization and identification functionalities for soldiers. For identification applications, each member of the team could transmit an LPI and LPD code. The soldier can then be identified using this code. The combination of secure communication with cooperative localization or short-range radar enables to have an Identification Friend or Foe (IFF). An automated and unmanned Ultra-Wideband (UWB) Perimeter Surveillance Sensor was designed to provide detection and tracking of personnel and vehicles at the perimeter of critical areas such as military installations and other facilities which was developed for low power dual band perimeter defense radar systems to operate alone or in concert with other sensors (Fig. 15.2).

Homeland Defense programs are receiving increasing attention which has brought to light a significant need to enhance existing technologies for sensor-based perimeter defense, particularly around military bases and high-risk commercial sites. Currently, most perimeter security systems are based on remote cameras or IR systems monitored by human operators; this approach is labor intensive and does not lend itself to easy automation.

Since the Federal Communication Commission (FCC) has declared UWB technology in 2002 [7], new communication standards and antenna designs have been proposed and a number of scientists and engineers has been attracted to this research area. The new communication standards and antenna designs have brought radar communication, military application, biomedical technology, and space communication through satellite into new era. To achieve the goal of UWB technology in different application areas, several designs of planar UWB antenna have been proposed. However, most of these antennas involve complex calculation and sophisticated fabrication process [8]. Therefore, a simpler method to design the UWB antenna having two triangular metal sheets with vertical slots has been proposed [9]. This sheet antenna has been designed for UWB and narrow pulsed communication systems. The antenna operates from 3 to 20 GHz. This antenna is suitable for use in UWB applications as it has the operating bandwidth of 13 GHz, i.e., a fractional and width of about 147%. This sheet antenna can be very useful for military application such as in radar communication, Unmanned Aerial Vehicles (UAV) communication, and designing surveillance sensor for critical military area.

The arrival of Ultra-Wideband technology for radar application allows the development of compact and relatively cheap sensors. Dedicated to military applications, radar devices using UWB are now a good tool of detecting obstacles for many applications. These sensors could be used to measure distances and positions with greater resolution than existing radar devices or to obtain images of objects buried underground or placed behind surfaces. Radar which exploits the radio frequency technique Ultra-Wide Band has been presented [10]. This study of UWB radars with short pulses is of great interest for electromagnetic detection and identification of targets at short and medium range, with a low cost and simple implementation. The radar, based on UWB Technology transmits very short electromagnetic pulses, which makes it possible to measure a very rich information transitory response of the target and to dissociate the various echoes at the reception. This brings great interest for obstacle detection and target identification in short-range application. UWB-based radar could be used in military application to detect target objects such as enemy armored vehicles, weapons, and marines.

The application of UWB technology to achieve the goal of Unmanned Ground Vehicles (UGV)/Unmanned Aerial Vehicle (UAV) for potential military application has contributed a lot. The relevance of UWB communications for ground-to ground or air-to-ground applications includes low probability of interception or detection (covertness), relative immunity to multipath effects, large spatial capacity compared to other wireless systems, simultaneous precision ranging and communicating, and extremely low transmit power which reduces the possibility of interfering with other radio systems. Outdoor UWB wireless network methodology is more relevant for

specifying realistic applications of UWB communications for unmanned system initiatives, which are an integral part of the Future Combat System. An assessment is made for specific areas such as UGV/UAV interoperability, UAV-to-UAV communications, and collision avoidance for small UAVs and the analysis presented in this paper [11] has been applied for determination of technology transition to evolving warfighting systems.

The high data rate capability of UWB systems for short distances has numerous applications for home networking and multimedia-rich communications in the form of Wireless Personal Area Networks (WPAN) and military applications. UWB systems could replace cables connecting camcorders, as well as other consumer electronics applications, such as laptops, DVDs, digital cameras, and portable HDTV monitors. In light of this assessment, research into countermeasures such as UWB radar, UWB Surveillance Sensors, UWB antennas, UWB radar imaging, IED detections, and others remains a key weapon in gaining advanced military technology during fight to protect warfighters from the very real threat posed by IEDs and hidden ordnance.

15.4 Conclusions

In this chapter, the application of UWB technology in military areas has been summarized. Because of its applicability to many advanced and existing technology, the topic of UWB has recently received significant attention from researchers and engineers. Its wider applications have been evaluated over several leading technology including military technology systems, government sectors, communication systems, surveillance systems, rescuing, logistics entertainment, public transport, and others. Engineers and defenses are planting their maximum effort to research on UWB for military technology to gain superiority over others. Many researchers and engineers are attracted to research on UWB technology characteristics and to equip military with the most recent technology. This chapter has covered most of the popular applications of UWB on military systems and their advantages, and shortcomings have been presented. Regarding the future research, to apply the knowledge of UWB technology, we need to know the physical characteristics, features, and how to make use of the features of UWB technology, how to correlate these features with the already existing technology like traditional narrowband frequencies. Thus, it will be easy to find new application areas of UWB technology and to conquer the limitations of the existing technology. Future research studies of UWB for military applications include:

- UWB Radar Technology for military and civilian applications, for rescue and disaster recovery.
- Imaging such as remote monitoring systems, UWB microwave imaging for detecting IEDs.

- Unmanned Ground Vehicles and Unmanned Arial Vehicles surveillance and communication monitoring.
- Satellite Systems and Satellite radio providers.
- Landmine detection using UWB radar.
- Vehicle-based and airborne detection systems.

References

1. Siwiak K (2001) Ultra-wide band radio: introducing a new technology, In: VTS 53rd IEEE vehicular technology conference, vol 2, pp 1088–1109
2. Kshetrimayum R (2009) An introduction to UWB communication systems. IEEE Potentials 28(2):9–13
3. Colson S, Hoff H (2005) Ultra-wideband technology for defence applications. In: IEEE international conference on ultra-wideband, pp 615–620, Sept 2005
4. Cramer R-M, Win M, Scholtz R (1998) Impulse radio multipath characteristics and diversity reception. IEEE Int Conf Commun 3:1650–1654
5. Immoreev I, Fedotov P (2002) Ultra wideband radar systems: advantages and disadvantages. In: IEEE conference on ultra wideband systems and technologies, pp 201–205, May 2002
6. Fontana R (2004) Recent system applications of short-pulse ultra-wideband (UWB) technology. IEEE Trans Microw Theory Tech 52(9):2087–2104
7. Lynch A, Genello B, Wicks C (2007) UWB perimeter surveillance. IEEE Aerosp Electron Syst Mag 22(1):8–10
8. Modi A, Gehani A (2012) Novel design of ultra wideband vertical slotted triangular (vst) sheet antenna. In: 5th international conference on computers and devices for communication (CODEC), pp 1–3, Dec 2012
9. Modi A (2013) Novel design of directive ultra wide band right angle triangular sheet antenna. In: International conference on microwave and photonics (ICMAP), pp 1–3, Dec 2013
10. Sakkila L, Rivenq A, Tatkeu C, El-Hillali Y, Ghys J-P, Rouvaen JM (2010) Methods of target recognition for UWB radar. In: IEEE intelligent vehicles symposium (IV), pp 949–954, June 2010
11. Levitt LJ, Fundamental communication range limitation of UWB communications for military application. Aviation and Missile Research, Development, and Engineering Center Redstone Arsenal, AL 35898-5000

Part III
Industrial Internet of Things

Recently, the technological advancement in Wireless Sensor Networks (WSNs), embedded systems, and low-power wireless communication has enabled numerous monitoring and control applications in different domains including Internet of Things (IoT) systems. An emerging class of IoT-enabled industrial production systems is called the Industrial Internet of Things (IIoT) that, when adopted successfully, provides huge efficacy and economic benefits to system installation, maintainability, reliability, scalability, and interoperability. Part III named Industrial Internet of Things mentions the state-of-the-art technologies along with accompanying challenges to realize such vision. Wide applications of IIoT are summarized in industrial domains. Specially, adopting such technology to the Physical Internet, an emerging logistics paradigm, is described in this part.

Chapter 16
An Overview on Industrial Internet of Things

16.1 Introduction

Recently, the technological advancement in identification technologies (e.g., RFID), Wireless Sensor Networks (WSNs), embedded systems and low-power wireless communication has enabled Internet of Things (IoT) to be deployed in numerous monitoring and control applications in different domains. An emerging class of IoT-enabled industrial production systems is called the Industrial Internet of Things (IIoT) that, when adopted successfully, provides huge efficacy and economic benefits to system installation, maintainability, reliability, scalability, and interoperability [1]. Although, IIoT applications are still in the early stage [2, 3], various important industries, which have been deployed such technology include environmental monitoring, healthcare service, inventory and production management, logistics, and supply chain systems [16].

The key component to create IIoT systems is communication technologies, which enable all devices, machines to connect, communicate, and exchange data together. In such a way, the system can monitor, collect, exchange, and analyze data, delivering valuable services that, in turn, enable the industry businesses to make more accurate and faster decisions.

16.2 Architecture of IIoT System

IIoT systems enable interconnecting a massive number of heterogeneous devices leveraging sensing, communication, and data processing technologies. Therefore, no single consensus on architecture for IIoT agreed universally leads to various proposed architectures depending on their specific application [4].

A reference architecture is still at the most basic model for IIoT architecture. Based on it, several multilayer models have been proposed to develop and adopt to

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_16

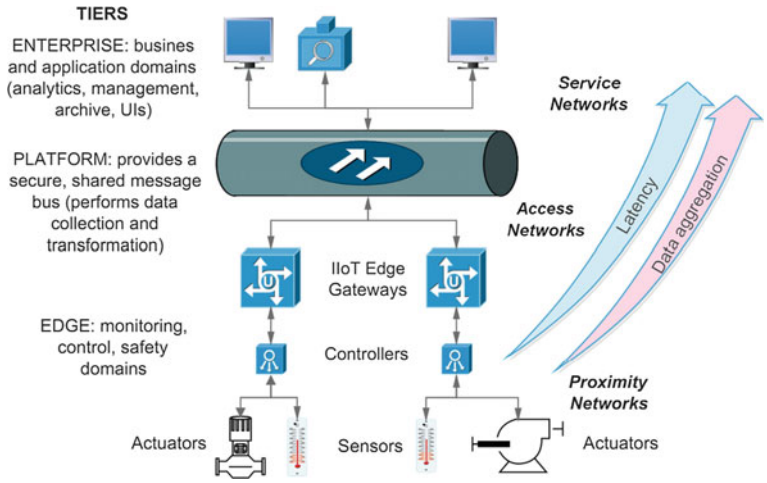


Fig. 16.1 The three-tier IIoT architecture [4]

provide specific IIoT services. For example, the International Telecommunication Union (ITU) defines five layers classified in the layer architecture of IIoT includes sensing, accessing, networking, middleware, and applications. The works in [5–7] suggested the identification of three major layers for IoT: perception layer (or sensing layer), network layer, and service layer (or application layer). Similarly, the work in [10] proposed a three-tier organizational structure for practical IIoT systems is proposed, which consists of three layers: a sensing layer, relay layer, and convergence layer. Such architecture was developed to support a real-time data routing in the IIoT systems. In a widely accepted IIoT architecture introduced in [15], three-layered networks were exploited to enable the connection of three industrial tiers: edge, platform, and enterprise tiers. Figure 16.1 illustrates the model in which three networks: proximity, access, and service are deployed and function.

Meanwhile, Liu et al. [8] designed an IoT application infrastructure based on the traditional OSI model that contains the physical layer, transport layer, middleware layer, and applications layer. In [9] a four-layered architecture derived from the perspective of offered functionalities includes four layers: the sensing layer, the networking layer, the service layer, and the interface layer.

Toward the Industry 4.0, the Reference Architecture Model Industry 4.0 (RAMI 4.0) [13] defines a 3D model as shown in Fig. 16.2 for the next-generation industrial manufacturing systems. This model includes three axes denoting the life cycle, value stream, and hierarchy levels. While the first two axes are related to the life cycle management of products, the third one is about the different component functionalities. This axis is also to describe the information technology representative with a communication layer embedded.

Similarly, in the context of Industry 4.0, all kinds of intelligent equipment (e.g., industrial robots) supported by wired or wireless networks are widely adopted, and

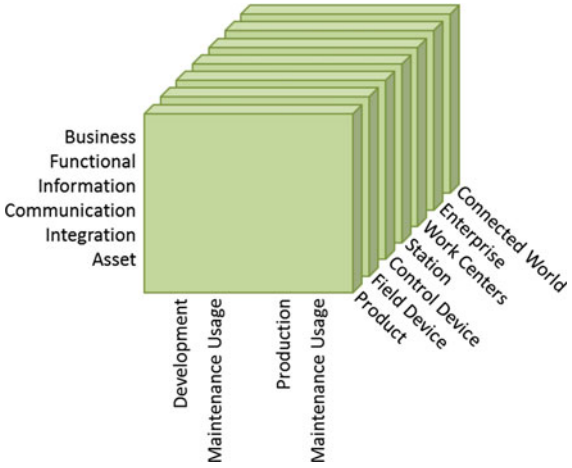


Fig. 16.2 The RAMI 4.0 cubic architecture. With the vertical axis representing software concerns, the horizontal axis representing life cycle stages, and the diagonal axis represents automation hierarchy [13]

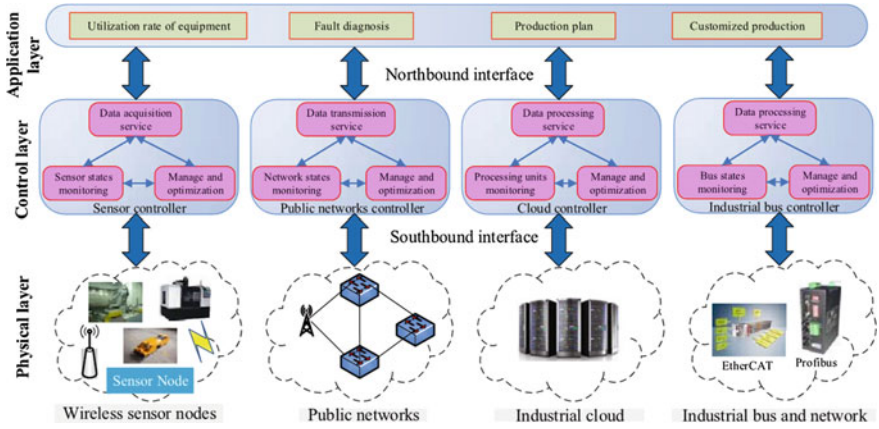


Fig. 16.3 Architecture of software-defined IIoT in the context of Industry 4.0

both real-time and delayed signals coexist. Based on the advancement of software-defined network technology, authors in [16] proposed a new IIoT architecture composed of three layers: the physical infrastructure, the control, and the application layers as described in Fig. 16.3. The architecture is to manage physical devices and provide an interface for information exchange

Recently, the Industrial Internet Consortium (IIT) [14] is working toward an interoperable IIoT architecture. In this mode, different viewpoints are incorporated such as formally business, usage, functional, and implementation views. Each aspect is modeled differently so as it can interact with the others. For example, the

implementation viewpoint is focused on the technologies and the system components that are required to implement the functionalities prescribed by the usage and functional viewpoints [4].

Referring to some IIoT scenarios in which a numerous number of things is moved dynamically, an adaptive architecture is needed to enable the IoT devices dynamically interact with other things in a real-time manner. In addition, the decentralized and heterogeneous nature of IIoT requires that the architecture provides IIoT efficient event-driven capability as well as on-demand services. In addition, the IIoT systems are difficult to implement and maintain at large scale due to the lack of an efficient, reliable, standardized, and low-cost architecture. Furthermore, the ever-changing demands of businesses and the vastly different needs of different end users should be met by providing customization functionality to the end users and organizations under a flexible SOA. Thus, an SOA is considered an efficient method achieve interoperability between heterogeneous devices in a multitude of way [10–12]. In addition, SOA is considered suitable for such demand-driven IoT services. In particular, SOA can integrate processes and information; and sharing such information can help create a better environment for real-time data exchange, real-time responsiveness, real-time collaboration, real-time synchronization, and real-time visibility across the entire industrial processes.

16.3 Key Enabling Technologies for IIoT

Realization of IIoT systems requires a set of technology-related components involving the IoT data generation, data processing and management, data communication. This section takes a brief introduction to overview the key technologies enabling the IIoT systems.

16.3.1 Identification Technology

Since the IIoT devices are connected to the Internet, the information relating to their physical status and context must be captured, coded, protected, and transferred accurately. In addition, users of IIoT device can be human, software applications, or other devices. Thus, to make a secure connection between devices or between devices with humans, identifying the devices and the users play an important role. The EPCglobal standard provides a solution that allows identifying all IoT objects uniquely. Particularly, the identification code is created in the line of RFID tag or barcode with the EPC standard. Thus, when the code is put into service, or networks, the information fed to an EPC Information Service (EPCIS) [22] is shared among IoT entities.

16.3.2 Sensor

Central to the functionality and utility of the IIoT systems are sensors, which are embedded in most of smart objects. Such sensors are capable of detecting events or monitoring changes in a specific quantity (e.g., pressure), communicating the event or change data to the cloud (directly or via a gateway) and, in some circumstances, receiving data back from the cloud (e.g., a control command) or communicating with other smart objects. Since 2012, thanks to the advancement of microelectromechanical systems (MEMS) sensors have generally shrunk in physical size and thus have caused the IIoT systems to mature rapidly. Concretely, low-cost, lower power, and small-sized sensors can be embedded into any smart objects to function [18]. Practically, industrial Wireless Sensor Networks are recognized to be a backbone in the most of IIoT systems enable monitoring and controlling in industrial processes and manufacturing [23].

16.3.3 Communication Technology

With respect to sending and receiving data, wired and wireless communication technologies enable the IoT devices to provide data connectivity. This has allowed the sensors embedded in smart objects to send and receive data over the cloud for collection, storage, and eventual analysis.

16.3.3.1 Wireless Technologies

The protocols for allowing IIoT devices to relay data wirelessly include wireless technologies such as RFID, NFC, Wi-Fi, Bluetooth, Bluetooth Low Energy (BLE), XBee, ZigBee, WirelessHART, ISA100.11a, as well as satellite connections and mobile networks using GSM, GPRS, 3G, LTE, or WiMAX [7, 9]. To support autonomous communication among machine or devices without human intervention, Machine-to-Machine (M2M) or Device-to-Device (D2D) communication technologies have been developed. Such autonomous networking technology is of paramount importance for several deployments of IIoT systems including smart manufacturing applications, healthcare systems, and home automation [21].

The flexibility and scalability of IIoT systems can be enabled thanks to the wireless communication technology, since it can support low power and long-range communication for the devices. However, the required Quality of Services (QoS) such as real time, reliability must be taken into account when designing and deploying the IIoT since the wireless communication is prone to error due to interference, noise, and collisions. To overcome such issues, some industrial wireless communication protocols have been proposed such as WirelessHART, or ISA100.11a.

16.3.3.2 Wired Technologies

Although wireless communication technologies bring great advantages for the IIoT system, wiring is still required in some IIoT solutions to function and operate safely and securely. Wired protocols employed by stationary intelligent devices include Ethernet, HomePlug [24], HomePNA, HomeGrid/G.hn [25], and LonWorks [26], as well as conventional telephone lines. Although innovative and advanced wireless technologies, especially in industrial applications are quickly allowing IoT to flourish, wired solutions will always be a part of IoT for years to come.

16.3.4 IIoT Data Management

The number of connected devices increases as well as raw data (unstructured data) that needs to be managed and processed. Thus, the term "big data" refers to these large data sets that need to be collected, stored, queried, analyzed, and generally managed in order to deliver on the promise of the IIoT. The big data technology relies on three metrics to describe and analyze the IoT data. They include volume (i.e., the amount of data they collect from their IoT sensors measured in gigabytes, terabytes, and petabytes), velocity (i.e., the speed at which data is collected from the sensor), and variety (i.e., the types of data collected for example, structured or unstructured data, video or picture data and so on) [17].

16.3.5 Cloud Computing

Cloud computing refers to an entity capable of accessing computing resources via the Internet rather than traditional systems where computing hardware is physically located on the premises of the user and any software applications are installed on such local hardware. Formally defined as in [19], "cloud computing" is "model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [19].

Cloud computing can support three important service models for the IoT: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). With such services, it allows any user with a browser and an Internet connection to transform smart object data into actionable intelligence. That is, cloud computing provides "the virtual infrastructure for utility computing integrating applications, monitoring devices, storage devices, analytics tools, visualization platforms, and client delivery to enable businesses and users to access IIoT-enabled applications on demand anytime, anyplace and anywhere" [20].

16.4 Major Application of IIoT

Success in deploying the IoT solution in different general domains [7, 11] promotes an increasing usage of IoT in industry. This section is to highlight applications of IIoT in some major industrial sectors.

16.4.1 Heath Care

IoT solutions are expected to provide promising health care services for the people thanks to the IoT's ubiquitous identification, sensing, and communication capability of health care systems [32–35]. Enabled by global connectivity, real time monitoring via connected devices can save lives in event of a medical emergency like heart failure, diabetes, asthma attacks, and so on. In addition, IoT healthcare device can collect medical and other required health data such as blood pressure, oxygen, and blood sugar levels, weight, and ECGs to transfer it to physicians for remote reporting and monitoring. These data are stored in the cloud and can be shared with an authorized person, who could be a physician, your insurance company, a participating health firm or an external consultant, to allow them to look at the collected data regardless of their place, time, or device.

16.4.2 Logistics and Supply Chain

With a huge number of logistics items are moved, handled, tracked, and stored by a variety of material handling equipment each day, the logistics and supply chain industry is a key player poised to benefit from the IoT. As more and more physical objects are capable of sensing, communicating, and data processing since they are equipped with ICT technologies such as bar codes, RFID tags, or wireless sensors, the IIoT enables seamless interconnection of the heterogeneous devices and to get complete operational visibility and allow for the best real-time decisions in the logistics and supply chain processes [27, 28]. Recently, an emerging paradigm of logistics called Physical Internet is expected to get a huge benefit from the IIoT solutions since it leverages the IoT technologies as its foundations [29–31].

16.4.3 Smart Cities

Practically, an increasing number of urban residents lead to more challenges of cities such as environmental deterioration, sanitation issues, traffic congestion, thwart urban crime. In such a context, IoT has the huge potential to tackle these issues by

developing the concept of smart cities. The fundamental of smart city is to rely on the smart technologies that collect the required data and analyze them for action in real time. Concretely, with advanced sensing and computation capabilities, data is gathered and evaluated in real time to extract the information, which is further converted to sable knowledge. This will enhance the decision making of city management and citizens [36, 37]. Basically, the benefits of smart cities are listed as follows:

- Traffic congestion control
  Smart cities provide safety and efficient solutions for traffic flows to avoid congestion issues. Such solutions exploit different types of sensors, as well as GPS data from driver's smartphones to determine the density, location, and speed of vehicles. In combination with the smart street light systems connected to the cloud management platform, the traffic flows are controlled and directed effectively. In addition, based on historical traffic data, smart solutions of smart city are able to predict the void location at which high density of traffic is passed. In such way, the potential congestion will be prevented.
- Smart Parking
  With the help of GPS data from drivers' smartphones (or road-surface sensors embedded in the ground on parking spots), smart parking solutions determine whether the parking spots are occupied or available and create a real-time parking map. When the closest parking spot becomes free, drivers receive a notification and use the map on their phone to find a parking spot faster and easier instead of blindly driving around [38].
- Environment Control and Management
  IoT-driven smart city solutions allow tracking parameters critical for a healthy environment in order to maintain them at an optimal level. For example, to monitor water quality, a city can deploy a network of sensors across the water grid and connect them to a cloud management platform. Sensors measure pH level, the amount of dissolved oxygen and dissolved ions. If leakage occurs and the chemical composition of water changes, the cloud platform triggers an output defined by the users. Another use case is monitoring air quality. For that, a network of sensors is deployed along busy roads and around plants. Sensors gather data on the amount of CO, nitrogen, and sulfur oxides, while the central cloud platform analyzes and visualizes sensor readings, so that platform users can view the map of air quality and use this data to point out areas where air pollution is critical and work out recommendations for citizens. Waste is an important factor causing the polluted environment. IoT-enabled smart city solutions help to optimize waste collecting schedules by tracking waste levels, as well as providing route optimization and operational analytics. Each waste container gets a sensor that gathers the data about the level of the waste in a container. Once it is close to a certain threshold, the waste management solution receives a sensor record, processes it, and sends a notification to a truck driver's mobile app. Thus, the truck driver empties a full container, avoiding emptying half-full ones.

- Smart Grids
  Smart girds refer to as IoT-based solutions to monitor and control the citizens' utilities including electricity, gas, and water. The solutions are based on smart meters and smart billing system to track and control the usage remotely.

16.5 Conclusions

The IIoT is widely considered to be one of the primary trends affecting industrial businesses today and in the future. Industries are pushing to modernize systems and equipment to meet new regulations, to keep up with increasing market speed and volatility, and to deal with disruptive technologies. Businesses that have embraced the IIoT have seen significant improvements to safety, efficiency, and profitability, and it is expected that this trend will continue as IIoT technologies are more widely adopted.

The IIoT solution greatly improves connectivity, efficiency, scalability, time savings, and cost savings for industrial organizations. It can unite the people and systems on the plant floor with those at the enterprise level. It can also allow enterprises to get the most value from their system without being constrained by technological and economic limitations. For these reasons and more, IIoT offers the ideal platform for bringing the power of the IIoT into both industrial sectors and enterprise.

References

1. Xu LD, He W, Li S (2014) Internet of things in industries: a survey. IEEE Trans Ind Inform 10(4):2233–2243
2. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun Surv Tutor 17(4):2347–2376
3. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
4. Sisinni E, Saifullah A, Han S, Jennehag U, Gidlund M (2018) Industrial internet of things: challenges, opportunities, and directions. IEEE Trans Ind Inf
5. Jia X, Feng Q, Fan T, Lei Q (2012) RFID technology and its applications in internet of things (iot). In: Proceedings of the 2nd international conference on consumer electronics, communications and networks (CECNet), pp 1282–1285
6. Domingo MC (2012) An overview of the internet of things for people with disabilities. J Netw Comput Appl 35(2):584–596
7. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
8. Liu CH, Yang B, Liu T (2014) Efficient naming, addressing and profile services in internet-of-things sensory environments. Ad Hoc Netw 18:85–101
9. Da Xu L, He W, Li S (2014) Internet of things in industries: a survey. IEEE Trans Ind Inf 10(4):2233–2243
10. Long NB, Tran-Dang H, Kim D (2018) Energy-aware real-time routing for large-scale industrial internet of things. IEEE Internet Things J 5(3):2190–2199

11. Miorandi D, Sicari S, De Pellegrini F, Chlamtac I (2012) Internet of things: vision, applications and research challenges. Ad Hoc Netw 10(7):1497–1516
12. Xu LD (2011) Enterprise systems: state-of-the-art and future trends. IEEE Trans Ind Inf 7(4):630–640
13. Flatt H, Schriegel S, Jasperneite J, Trsek H, Adamczyk H (2016) Analysis of the cyber-security of industry 4.0 technologies based on RAMI 4.0 and identification of requirements. In: IEEE 21st international conference on emerging technology and factory automation, pp 1–4
14. Industrial internet reference architecture. http://www.iiconsortium.org/IIRA.htm
15. IoT 2020 (2016) Smart and Secure IoT Platform. International electro-technical commission
16. Wan J et al (2016) Software-defined industrial internet of things in the context of industry. IEEE Sens J 16(20):7373–7380
17. Sezer OB, Dogdu E, Ozbayoglu AM (2018) Context-aware computing, learning, and big data in internet of things: a survey. IEEE Internet Things J 5(1):1–27
18. Combaneyre F (2015) Understanding data streams in IoT, SAS White Paper
19. National Institute of Standards and Technology (U.S. Dept. of Commerce) (2011) The NIST definition of cloud computing, Special Publication, pp 800–145
20. Canellos D (2013) How the "Internet of Things" will feed cloud computing's next evolution. Cloud Security Alliance Blog, 5 June 2013)
21. Montori F, Bedogni L, Di Felice M, Bononi L (2018) Machine-to-machine wireless communication technologies for the internet of things: taxonomy, comparison and open issues. Pervasive Mobile Comput 50:56–81. ISSN 1574-1192
22. EPCIS, GS1 Standard. https://www.gs1.org/epcis/epcis/1-1
23. Sheng Z, Mahapatra C, Zhu C, Leung VCM (2015) Recent advances in industrial wireless sensor networks toward efficient management in IoT. IEEE Access 3:622–637
24. HomePlug. http://www.homeplug.org/
25. HomePNA. http://www.homepna.org/
26. LonWorkds Technology. http://www.lonmark.org/news_events/press/2008/1208_iso_standard
27. Karakostas B (2013) A DNS architecture for the internet of things: a case study in transport logistics. Procedia Comput Sci 19:594–601
28. Sun C (2012) Application of RFID technology for logistics on internet of things. AASRI Procedia 1:106–111
29. Montreuil B (2011) Toward a physical internet: meeting the global logistics sustainability grand challenge. Logist Res 3(2–3):71–87
30. Montreuil B, Meller RD, Ballot E (2012) Physical internet foundations. IFAC Proc 45(6):26–30
31. Tran-Dang H, Kim D (2018) An information framework for internet of things services in physical internet. IEEE Access 6:43967–43977
32. Domingo MC (2012) An overview of the internet of things for people with disabilities. J Netw Comput Appl 35(2):584–596
33. Alemdar H, Ersoy C (2010) Wireless sensor networks for healthcare: a survey. Comput Netw 54(15):2688–2710
34. Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak K (2015) The internet of things for health care: a comprehensive survey. IEEE Access 3:678–708
35. Sundaravadivel P, Kougianos E, Mohanty SP, Ganapathiraju MK (2018) Everything you wanted to know about smart health care: evaluating the different technologies and components of the internet of things for better health. IEEE Consum Electron Mag 7(1):18–28
36. Ahlgren B, Hidell M, Ngai EC (2016) Internet of things for smart cities: interoperability and open data. IEEE Internet Comput 20(6):52–56
37. Jin J, Gubbi J, Marusic S, Palaniswami M (2014) An information framework for creating a smart city through internet of things. IEEE Internet Things J 1(2):112–121
38. Lin T, Rivano H, Le Mouël F (2017) A survey of smart parking solutions. IEEE Trans Intell Transp Syst 18(12):3229–3253

Chapter 17
Energy-Aware Real-Time Routing for Large-Scale Industrial Internet of Things

17.1 Introduction

Recently, the technological advancement in wireless sensor networks (WSNs), embedded systems and low-power wireless communication has enabled numerous monitoring and control applications in different domains including Internet of Things (IoT) systems [1]. An emerging class of IoT-enabled industrial production systems is called the Industrial Internet of Things (IIoT) that, when adopted successfully, provides huge efficacy and economic benefits to system installation, maintainability, reliability, scalability, and interoperability. However, high energy consumption of IIoT devices remains a significant challenge, which could prevent the IIoT perspective from becoming a reality.

Generally, industrial Wireless Sensor Networks (IWSNs) play a significant role as the backbone of IIoT systems. Thus, since the data collection relies mainly on these sensors and smart devices, the IWSNs become a dominant source of energy consumption in the whole systems. In order to obtain the power saving, optimizing of sensing, processing, and communication is an effective solution [2]. For instance, by opportunistically offloading data of the IIoT devices to smart devices being carried by the workforce in the factory settings, a heuristic and opportunistic link selection algorithm (HOLA) introduced in [3] achieved the energy efficiency. In addition, these smart devices equipped with multiple radio links such as Bluetooth, Wi-Fi, and 3G/4G LTE could select independently the best link to transmit the data to the Cloud based on the quality and energy cost of the link. To sum up, HOLA reduced the overall energy consumption of the IIoT network effectively as well as balanced it across all devices of the network.

Designing energy-efficient deployment schemes is another way to overcome the challenge. Typically, there are four types of topological structures exploited for the deployment of large-scale IWSNs: mesh, plane, hierarchical, and hybrid. However, unlike IWSNs and the traditional Wireless Sensor Networks (WSNs), in [4–6], the authors showed that IIoT could be applied to a larger scale and become more

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_17

complex. In addition, IIoT applications have more sophisticated and heterogeneous requirements than classic ad hoc wireless and sensor network applications. To sum up, to develop the IIoT technology for harsh environment, such as industrial firms, effectively becomes demanding and challenging for researchers because it has been impossible to transplant all schemes deployed in WSNs or even large-scale IWSNs in the IIoT directly. In other words, novel deployment models must be designed and validated to be deployed appropriately in the IIoT systems. Recently, one of these schemes proposed in [7] structured the IIoT systems into three layers: sense layer, gateway layer, and control layer. The fundamental purpose of such an architecture is to distinguish nodes as sense nodes, gateway nodes, and control nodes, thus the traffic loads can be balanced, and thus network lifetimes may be prolonged.

Clustering proved to be an efficient technique to solve the problem of power consumption in the large-scale I/WSNs can be employed in the large-scale IIoT system to obtain the same target. On the basis of clustering method, the homogeneous devices of the system are grouped into groups called clusters. Each cluster is managed by a Cluster Head (CH), which in turn collects or aggregates the data of cluster members and then forwards to the final destination or other cluster heads. In addition, CH role is rotated fairly to all the cluster members over predefined rounds. In this way, the systems can reduce the overall energy consumption and balance it across the nodes. However, depending on specific conditions (i.e., network topology, network channel conditions, node status, etc.), efficient clusters are formed by different ways. For example, in [8], the researchers proposed the clustering algorithm which would search intermediate nodes (IN) to manage subclusters while a CH aggregated data from INs. As a result, it could not only optimize energy but also prolong the network lifetime in IWSNs. Meanwhile, INs were considered relays between neighboring CHs, as introduced in [9]. By using both feedback and non-feedback fountain-code cooperative communication, the performance of data transmission improved significantly among neighboring CHs. Moreover, through an analysis of the effect of relay cluster size selection on the end-to-end data rate in multi-hop cooperative relaying performance in [10], the authors demonstrated that the selection of the optimal cluster size plays a vital part in maximizing the relay throughput and the end-to-end data rate. In addition, to address the issue of imbalanced channels from relays to a source/a destination in dual-hop cooperation technique in clustered wireless networks, based on the Three-Stage Relaying (TSR) framework, two heuristic algorithms in [11] were designed to achieve the trade-off between optimality and computational complexity.

In terms of a protocol stack for IIoT applications, several low-power wireless technologies such as ZigBee, 6LoWPAN, WirelessHART, IAS100.1 can be employed in field devices (i.e., sensors, actuators, etc.) to deal with the energy consumption issues. Palattella et al. [12] listed three core requirements, including a low-power communication stack, a highly reliable communication stack, and an Internet-enabled communication stack. As a result, to satisfy these requirements, the authors demonstrated that IEEE 802.15.4 is the standard with the longest standing impact. A low-power physical layer (PHY) upon which most IIoT technologies have been built was denoted by IEEE 802.15.4. To demonstrate the vital role of IEEE 802.15.4 in the industrial environment, Al Agha et al. in [13] reviewed briefly specifications of

all IEEE 802.15.4-based wireless technologies employed in IWSNs, and then they relied on IEEE 802.15.4 to develop the wireless sensor communication module running an industrial ad hoc mesh networking protocol using IEEE 802.15.4 PHY layer. Consequently, their IEEE 802.15.4-based protocol could satisfy criteria of the harsh environment which was time constraints, energy consumption routing strategy, and human walking-speed mobility.

However, toward the practical deployment, IWSN solutions should be versatile, simple to use and install, long lifetime and low-cost devices. Indeed, the conflict of requirements leads to the difficulty of their combination met. Thus, depending on specific applications, the systems should be designed properly. For example, the key applications of IWSNs in industrial production include three groups: environmental sensing, condition monitoring, and process automation, which require increasing critical level (e.g., time-critical transmission, priority-critical data transmission) [14]. The scope of our research work is then to design the IIoT system with IWSNs as backbone supporting the lowest level of industrial requirements, i.e., non-time-critical application. To support such an industrial application, this chapter proposes a hierarchical framework for the of IIoT systems, which consists of three types of elements: I/O devices, routers, and gateways. In order to achieve the overall energy efficiency of the systems, the lower power stack technology (i.e., IEEE 802.15.4a with CSMA/CA MAC protocol) is employed in all I/O devices. Clustering method is used to group these devices and routers in clusters by an optimal clustering algorithm taking into account the energy consumption. These clusters are managed by routers selected as CHs appropriately. These CHs then forward gathered data on the lower energy consumption route. To sum up, the main contributions of this chapter are as follows:

- A three-tier organizational structure for practical IIoT systems is proposed, which consists of three layers: a sensing layer, relay layer, and convergence layer.
- In the sensing layer and relay layer including IO devices and routers, we design cluster frameworks to organize an optimal energy cluster structure. I/O devices would calculate a transmission distance and a potential index (PI) for each router, considered as a CH, using the residual energy and position information transmitted from routers. Consequently, I/O devices only join the cluster whose router has the greatest potential index.
- Based on the aforementioned architecture, we develop a routing algorithm allowing the CHs in the network to forward their data in the lowest energy consumption route. Taking into account the time-critical requirement of the industrial environment, the selected routing path exploiting the shortest hop count from the CHs to gateways allows systems to obtain low end-to-end delay.
- The enhancement of the IIoT systems when deploying the proposed scheme is examined by simulations result, which indicated threefold of achievement: saving energy, increasing network lifetime and lowering end-to-end delay.

The rest of this chapter is organized as follows. Section 15.2 summarizes the related works. Section 15.3 describes the system model in the large-scale IIoT. The

proposed scheme, namely Energy-aware Real-time Routing Scheme (ERRS), in this chapter are presented in Sect. 15.4, followed by a performance evaluation in Sect. 15.5. Finally, the conclusion and future work are presented in Sect. 15.5.

17.2 Related Works

In this section, we present some related and well-known routing protocols recently proposed for IIoT systems.

One of the purposes of designed IoT solutions is to improve the reliability of IIoT systems through advanced monitoring applications since the IoT provides real-time information acquired from connected devices and their interactions, as well as perform real experimentation through the extensive testbeds deployed in industrial environments. For instance, in [15], Civerchia et al. adopted IoT protocols to develop an advanced IIoT solution aiming for a new generation of smart factories. Their proposed scheme, namely NGS-PlantOne solution, was designed to support advanced predictive maintenance applications in the industrial environments, which was based on four main components: sensors devices, gateway, remote control and service room, and open platform communications. By evaluating the proposed scheme through a real testbed over a period of 2 months, the proposed IIoT solution could guarantee that all nodes could either communicate each other with an acceptable delay or reduce battery power consumption significantly. Meanwhile, to deploy an IoT network for a smart factory, the authors [16] employed IPv6 over Low-power Wireless Personal Area Network (6LoWPAN) in combination with the IPv6 Routing Protocol for Low power and Lossy Networks (RPL) and Constrained Application Protocol (CoAP) for IWSNs. In the industrial network topology, the authors designed a wired plant automation network which can not only be connected directly to the Internet but also communicate with an IWSN. The simulation and testbed experiments demonstrated that all topologies measured have lower average latency than 400 ms.

For large-scale systems, clustering is one of the most efficient methods to enhance the overall efficiency of the systems. Depending on the structure as well as characteristics of networks, clusters are formed under different specific conditions. For example, based on local-world theory, a new clustering weighted evolving model of WSNs was proposed for IoT in [17]. The proposed scheme took sensor energy, transmission distance, and flow into account to determine edge weight and vertex strength. Due to controlling topology growth and weight dynamics, the schemes formed the uneven clusters prevented an energy hole and maintain a balance in the energy consumption of the entire IoT network. Meanwhile, in [18], a predictive energy consumption efficiency (PECE)-based clustering routing algorithm was divided into two stages, which were cluster formation and stable data transfer. In the first stage, to determine a cluster head, the algorithm considered degree and relative distances between nodes. Next, in the second stage, Bee Colony Optimization (BCO) was used to predict the route yield of each routing part from the source node to the sink node. It leads to either improving the overall network performance or prolonging the

survival time of the network. Whereas [19] selected the forward routes which could connect more multi-cast receivers using the offset back-off scheme and broadcast characteristic of wireless communication, [20] presented an energy-balanced routing method based on Forward-Aware Factor (FAF-EBRM) for WSNs. Thanks to an awareness of link weight and forward energy density, the FAF-EBRM can denote the reliable next-hop node in routes. Then, based on a spontaneous reconstruction mechanism for local topology, the FAF-EBRM outperformed the conventional schemes, such as Low-Energy Adaptive Clustering Hierarchy (LEACH) and Energy-Efficient Uneven Clustering (EEUC), in terms of balancing energy consumption, expanding network lifetime, and guaranteeing high Quality of Service (QoS) of WSN. An energy-efficient cluster-based routing protocol, called Quasi Group Routing Protocol was proposed in [21] for multi-hop smartphone networks based on Wi-Fi Direct. Accordingly, all devices in the network were grouped into quasi-groups (clusters) and cluster heads were selected based on three different approaches: (1) the device with the highest ID in the surroundings, (2) the peer that has the shortest average distance from the other nodes, (3) the node with less mobility with respect to its neighbors. In addition, the algorithm allowed the devices to rotate the CH responsibility and efficiently manage the consumption of their batteries. The proposed scheme validated through simulation could save a significant amount of energy.

Generally, the energy consumption in WSNs is impacted by not only the total amount of data sensed and collected, but also the routing schemes. Although collaborative charging method can effectively deal with the influences, the utility of Wireless Rechargeable Sensor Networks (WRSNs) is not optimized since the excessive harvested energy cannot be allocated for sensing, transmitting data properly. Zhang et al. [22] developed the BEAS algorithm to efficiently manage the battery energy usage, and the DSR2C algorithm for obtaining the optimal sensing rate and routing. Accordingly, the energy consumption rate was under control, saying that was kept to be lower than the energy allocation. In addition, sensor nodes choose a path with the lowest energy consumption to transfer data by using a charging-aware routing protocol. In this way, energy consumption of the network can be balanced with lifetime prolonged as well. Although the energy allocation problem is formulated based on the energy harvesting rate and current battery level, it is heuristically solved without considering the global optimality. With regards to routing the information in emergency and public safety networks, a new multi-hop routing protocol called Optimized Routing Approach for Critical and Emergency Networks (ORACE-Net) was presented in [23]. As a metric-based routing protocol, the proposed routing scheme relied on real-time end-to-end link quality estimation metrics including end-to-end signal strength and end-to-end hop count for routing decisions. Most notably, since the mobility of nodes was taken into account in the metric estimation, ORACE-Net protocol could be realized in reliable emergency systems. Investigated by simulation processes, the proposed protocol outperforms the other studied protocols in terms of packet reception rate and energy consumption. In addition, it increased the body-to-body network lifetime and reliability.

17.3 System Model

17.3.1 Network Topology

Typically, it is feasible to deploy dynamic routing protocols for the small-scale WSNs, however a large amount of energy is spent on data process and data communication. To overcome the disadvantage in our large-scale networks, a static routing method designed based on the tiered structure of network topology is used since it could migrate traffic load effectively [24, 25]. Accordingly, the network model is constructed from three layers, namely, sensing layer, relay layer, and convergence layer as depicted in Fig. 17.1. The sensing layer contains industrial sensors and radio-frequency identifications (RFIDs), called I/O devices, which acquire and transmit data to the relay layer. In the relay layer, routers are interconnected and forward data to a gateway responsible for uploading data to the Internet. According to the functions and specifications of I/O devices, routers, and gateways, the communication rule between two IIoT elements is defined as follows:

- An I/O device, s_i, cannot directly communicate with another I/O devices ($s_i, s_j \in S; d_{s_i s_j} \leq d_{s_{\max}}$);
- An I/O device, s_i, can only transmit data packets to a router, $r (s \in S, r \in R; d_{sr} \leq d_{s_{\max}})$;
- A router, r_i, can reach another, r_j, or a gateway, $g (r_i, r_j \in R, g \in G; d_{r_i r s_j} \leq d_{s_{\max}}, d_{r_i g} \leq d_{r_{\max}})$.

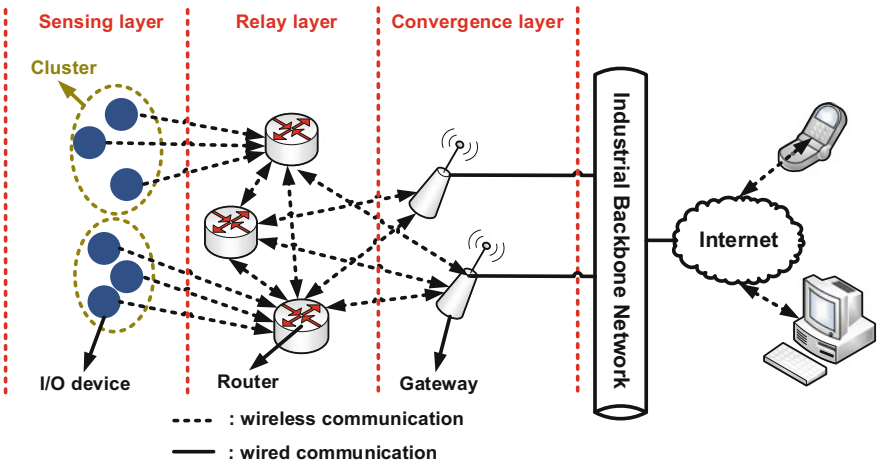


Fig. 17.1 Network topology for large-scale IIoT deployment in the industrial environment

Table 17.1 System model notation

Symbol	Definition
$d_{u,v}$	The distance between node u and node v
$\mathcal{S}, \mathcal{R}, \mathcal{G}$	The set of I/O devices, routers, and gateways, respectively
$s \in \mathcal{S}, r \in \mathcal{R}, g \in \mathcal{G}$	I/O device, router, and gateway, respectively
$d_{s\max}, d_{r\max}$	The communication distance of I/O devices, and routers, respectively
E_{elec}	The energy consumed to process data
$\epsilon_{\text{mp}}, \epsilon_{\text{fs}}$	The coefficients of distance-dependent multi-path fading and free-space channel models
d_0	The threshold for switching to the multi-path fading channel
E_{DA}	The energy for aggregation and compression
E_s, E_r, E_g	The energy consumption of an I/O device, a router, and a gateway, respectively
$E_{tx}^{ij}(k, d_{ij})$	The energy consumption of an I/O device i when it sends a k-bit packet over a distance d_{ij} to a device j
$E_{rx}^{ij}(k)$	The energy consumption of an I/O device j when it receives a k-bit packet from an device i
E_{init}	The initial energy of IIoT elements

17.3.2 Variable Definition

Table 17.1 lists and defines all notations of parameters and variables used in this chapter.

17.3.3 Energy Model

To evaluate the energy efficiency of ERRS, the chapter used the energy model introduced in [26]. The energy for transmitting k-bits long packet from node u to node v is computed by

$$E_{tx}^{uv}(k, d_{u,v}) = \begin{cases} E_{\text{elec}} \times k + \epsilon_{\text{fs}} \times k \times d_{uv}^2 & \text{if } \mathrm{d}_{uv} \leq d_0 \\ E_{\text{elec}} \times k + \epsilon_{\text{mp}} \times k \times d_{uv}^4 & \text{if } \mathrm{d}_{uv} \geq d_0 \end{cases}. \tag{17.1}$$

To receive a k-bit packet, the amount of energy consumed at node v is given by

$$E_{rx}^{uv}(k) = (E_{\text{elec}} + E_{DA}) \times k. \tag{17.2}$$

Considering a cluster with a set of I/O devices N ($|N| < |S|$), E_s, E_r, and E_g are, respectively computed by

$$
\begin{aligned}
E_s &= E_{tx}^{sr}(k, d_{sr}), \\
E_r &= \sum_{i=1}^{|N|} E_{rx}^{s_i r}(k) + \sum_{z,j=1}^{|M|} \left\{ E_{tx}^{rr_j}(k, d_{rr_j}) + E_{rx}^{r_z r}(k) \right\} \\
&\quad + \sum_{u=1}^{|O|} E_{tx}^{rg_u}(k, d_{rg_u}), \\
E_g &= \sum_{j=1}^{|P|} E_{rx}^{r_j g}(k).
\end{aligned} \tag{17.3}
$$

where M, O is a set of neighbor nodes of router r, and P is a set of neighbor nodes of gateway g ($|M|, |P| \leq |R|, |O| \leq |G|$).

Therefore, the energy consumption $E_{\text{cluster},i}$ in a cluster i is calculated by

$$
E_{\text{cluster},i} = \sum_{j=1}^{|N|} E_{s_j} + E_{r_i}. \tag{17.4}
$$

The total energy consumption in a round E_{round} is given by

$$
E_{\text{round}} = \begin{cases} E_{\text{cluster},i} + E_g & \text{if } L = 1 \\ E_{\text{cluster},i} + \sum_{j=1}^{L} E_{r_j} + E_g & \text{if } L = 1 \end{cases}, \tag{17.5}
$$

where the round is defined as the data aggregated from an I/O devices to a gateway, and L is the number of hops from cluster i to gateway g.

Since the main purpose that is to minimize energy consumption for IIoT systems, the optimization formulation of this chapter is determined by

$$
\begin{aligned}
&\textbf{Minimize} && E_{\text{round}} \\
&\textbf{subject to} && E_s, E_r, E_g \geq 0 \\
&&& E_{\text{cluster}} \geq 0 \\
&&& |N| \leq |\mathcal{S}| \\
&&& |M|, |P| \leq |\mathcal{R}| \\
&&& |O| \leq |\mathcal{G}|
\end{aligned} . \tag{17.6}
$$

According to the optimization model, to minimize the entire energy consumption of a large-scale IIoT system, first, the power consumed for one round should be optimized, which leads to a demand for minimizing the energy consumption for either transmission or reception of I/O devices , routers, and gateways. Next, the

routing algorithm would search the lowest energy consumption path that owns the least number of hops from a CH to a gateway. Moreover, in harsh environments, such as IIoT, the hop counts also play a vital part in forwarding data in real time. In Sect. 17.4, the routing process would be described in detail.

17.4 Energy-Aware Real-Time Routing Scheme (ERRS)

In the network, to update the information of neighbor nodes before creating a cluster or routing a data, IIoT elements in the chapter rely on the proactive protocol like [27]. Routers and gateways periodically broadcast advertising packets (ADV) after a predefined period of time. The ADV packet contains information about position, hop counts, and residual energy of all IIoT elements in the network, as depicted in Fig. 17.2. After receiving the ADV packets, all IIoT nodes can take advantage of available information to select the cluster which owns a potential router, and construct the data route to gateways.

17.4.1 Clustering Scheme

Because the cluster organization is relied entirely on computing the energy efficiency required and deciding which potential cluster an I/O device should join, the network lifetime in the IIoT system would be extended significantly. The formation of a cluster depending on the number of routers n ($n = 1$ or $n > 1$) in the communication range of I/O devices is performed through the procedure of packet exchange between I/O

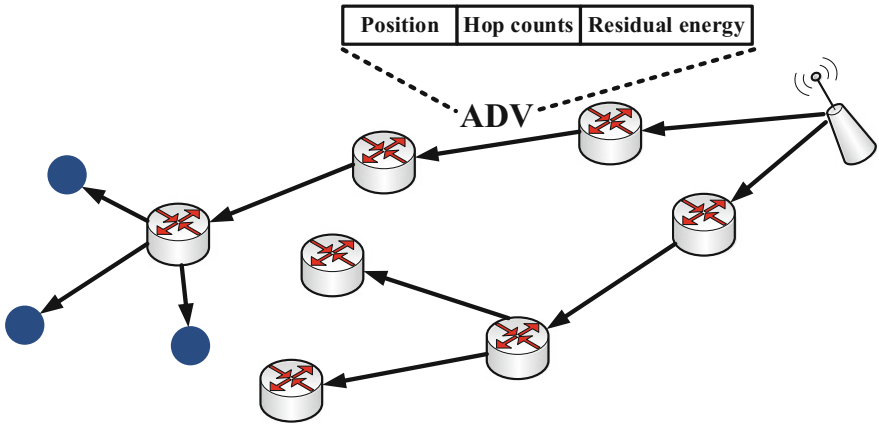


Fig. 17.2 Update information from ADV packets

Table 17.2 List of packets used by I/O devices and routers in the cluster formation process

n	Packets	Meanings
$n=1$	HELLO-REQ	Start cluster formation signaled by routers
	REV-HELLO-REQ	Receipt of HELLO-REG informed by I/O devices
	YES-RES	Allowance of cluster joining confirmed by routers
$n>1$	HELLO-REQ	Start cluster formation signaled by routers
	WAIT-RES	Receipt of HELLO-REG informed by I/O devices
	JOI-REQ	Cluster joining requested by routers
	SEL-RES	Selected clusters with CHs informed by I/O devices

devices and routers. For clarity description of the clustering process, the important packets are listed in Table 17.2.

(1) $n=1$: As illustrated in Fig. 17.3a, to initialize the clustering process, an HELLO-REQ (hello-request) packet is propagated from router r_D its neighbor I/O devices including s_A, s_B, and s_C to inform them of its role. Unlike ADV packet, HELLO-REG is as a beacon message without included information aiming to inform neighbor I/O devices that the cluster formation process starts. Due to receiving only one HELLO-REQ packet sent from r_D, the three I/O devices immediately decide to be cluster members of the cluster of router r_D without considering the energy efficiency. When s_A, s_B, and s_C expect to join the cluster of r_D, they transmit REV-HELLO-RES (receive-hello-response) packets containing an identification number, residual energy, and position information to acknowledge receipt of the HELLO-REQ packet and to get permission from CH r_D. If the REV-HELLO-RES is received successfully by CH r_D, r_D allows the three I/O devices to be its cluster members by sending YES-RES (yes-response) packets including the routing information with CH to s_A, s_B, and s_C. Then, the I/O devices update their own routing table, and the cluster organization is completed, as described in Fig. 17.4.

(2) $n>1$: Unlike I/O devices s_B, and s_C in Fig. 17.3b, s_A can communicate with routers r_D, r_E, and r_F; therefore, so we consider only the cluster selection of I/O device s_A. To inform three routers of the successful receipt of HELLO-REQ packets from them, I/O device s_A broadcasts WAIT-RES (waiting-response) packets containing its residual energy and position information.

Including the residual energy of s_A, the average residual energy of all I/O devices in each cluster calculated by the router is given by

$$\bar{E}_{\text{cluster-res}_j} = \frac{\sum_{i=1}^{|W|} E_{s\text{-res}_i}}{|W|}, \tag{17.7}$$

where $E_{s\text{-res}_i} = E_{\text{init}} - \sum_{\text{round}=1}^{\infty} E_{s_i}$ is the residual energy of an ith I/O device, and W is a set of I/O devices in a cluster. Then, the JOI-REQ (joining-request) packet

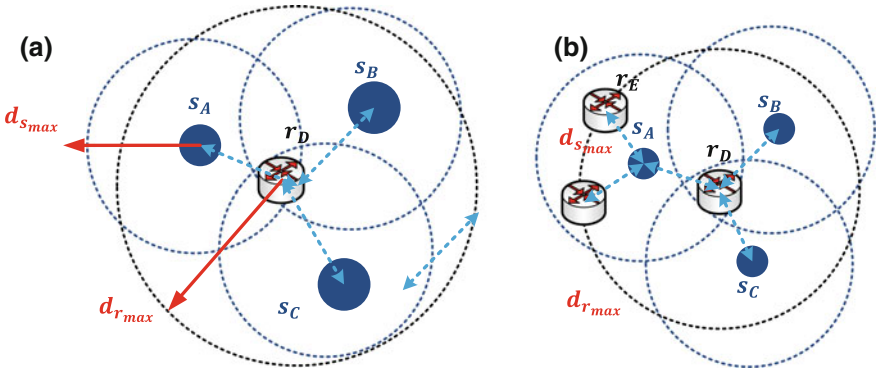


Fig. 17.3 Router distribution: **a** one router in the communication range of each I/O device, and **b** three routers in the communication range of one I/O device

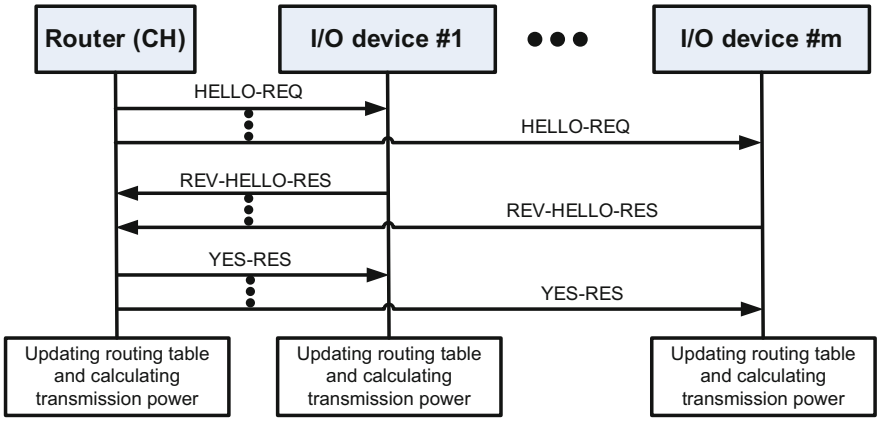


Fig. 17.4 Creating a cluster with $n = 1$

containing the average residual energy and location of r_D, r_E, and r_F is replied from the three routers to s_A.

By employing the information in JOI-REQ, the PI that s_A computes for each router is obtained from the following:

$$\mathrm{PI}_{s_k r_j} = \frac{E_{s\text{-res}_k}}{\bar{E}_{\text{cluster-res}_j}} \times \frac{1}{d_{s_k r_j}}, \tag{17.8}$$

where $d_{s_k r_j} = \sqrt{\left(x_{s_k} - x_{r_j}\right)^2 + \left(y_{s_k} - y_{r_j}\right)^2}$. Among three routers, s_A decides to join the cluster whose the CH owns the greatest PI and informs r_D, r_E, and r_F of its selection through a SEL-RES (selection-response) packet. The router, as the potential CH, is defined as

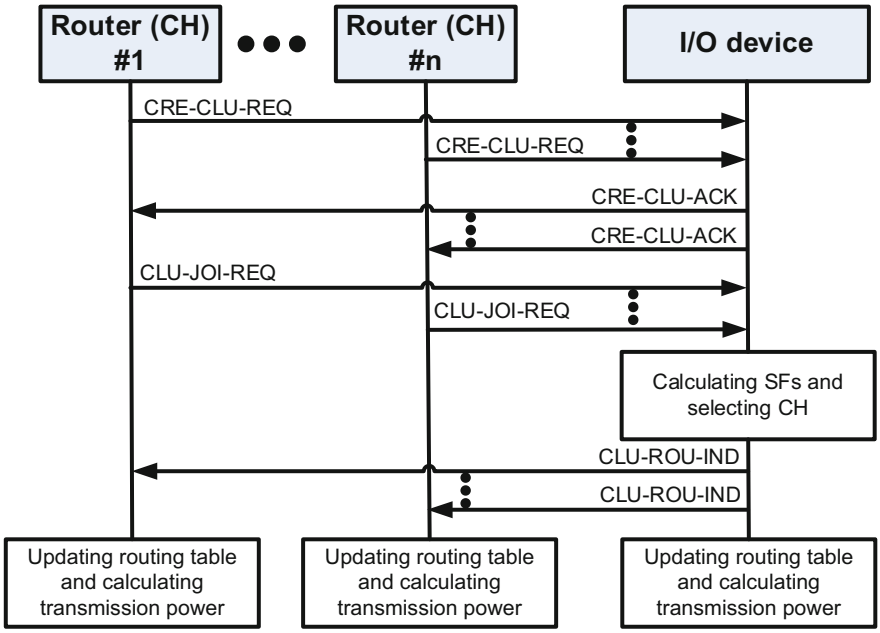


Fig. 17.5 Creating a cluster with $n > 1$

$$\mathrm{CH} = \left\{ r_j \,\middle|\, \arg\min\left\{ \mathrm{PI}_{s_k r_j} \right\}, \quad r_j \in \mathcal{R}, \quad s_k \in \mathcal{S} \right\}. \tag{17.9}$$

The process of creating a cluster for this case is summarized in Fig. 17.5.

17.4.2 Routing Scheme

After finishing the cluster organization process, the IIoT system includes formed clusters and their selected CHs (selected routers) as illustrated in Fig. 17.6.

In a cluster, after aggregating the data transmitted by I/O devices during their allocated transmission time, each router provides an energy-aware real-time routing function to forward the data to gateways. At the beginning of routing data packets to a gateway, an INI-ROU (initializing-route) message is propagated from a router to its neighbors within the communication range $d_{r_{\max}}$. If the neighbors receive the INI-ROU successfully, they request permission to forward the data by transmitting their WAIT-ACK (waiting-acknowledgment) message consisting of their address, hop count, residual energy, and positional information.

$$g_{\mathrm{des}} = \left\{ g_i \,\middle|\, \arg\min_{Z} \{ d_{r_A g_i} \}, \quad g_i \in Z, \quad Z \subset \mathcal{G} \right\}, \tag{17.10}$$

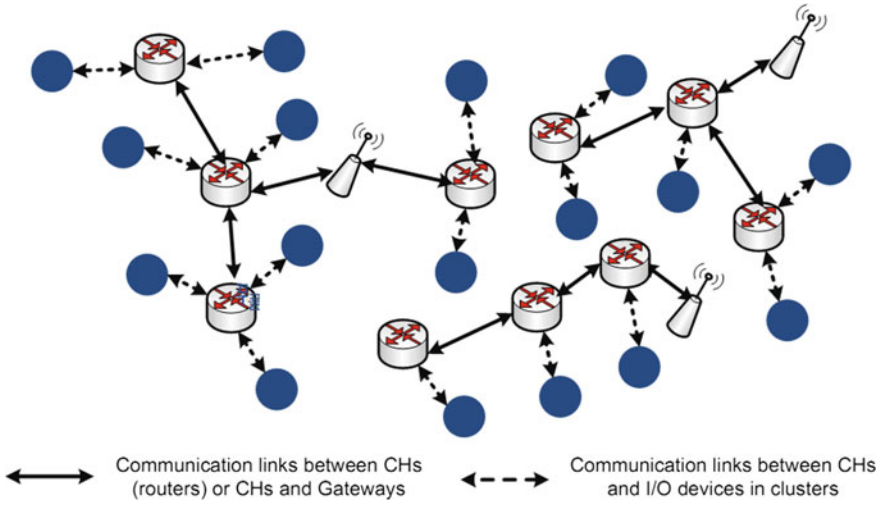


Fig. 17.6 A sample topology of IIoT system with clusters and CHs (routers) selected after finishing one round of clustering organization

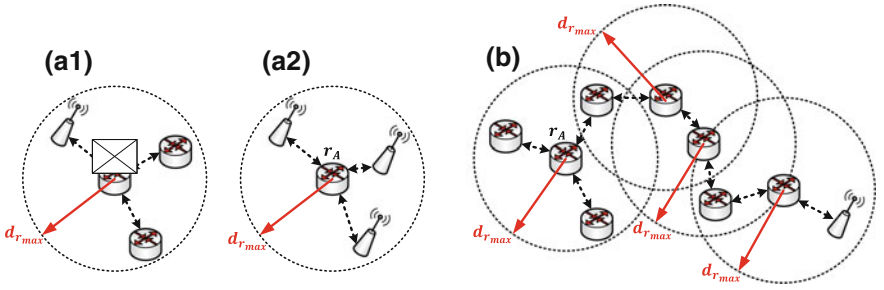


Fig. 17.7 Gateway distribution: **a** the gateways are distributed in r_A's communication range, and **b** the gateways are distributed out of r_A's communication range

where Z is a set of gateways in r_A's communication range, and $d_{r_A g_i} = \sqrt{\left(x_{r_A} - x_{g_i}\right)^2 + \left(y_{r_A} - y_{g_i}\right)^2}$.

In the case that there is only a gateway in r_A's communication range like Fig. 17.7a1, the data is forwarded directly from r_A to the gateway directly after r_A receives WAIT-ACK from the gateway. However, regarding Fig. 17.7a2, if several neighbor gateways are distributed in the communication range of r_A, the routing algorithm employed the transmission distance to select the gateway. To complete the route, r_A only sends data to the closest gateway g_{des} determined by

$$\bar{d}_{r_A r} = \frac{\sum_{i=1}^{|Q|} d_{r_A r_i}}{|Q|} \tag{17.11}$$

where Q is a set of r_A's neighbor nodes and $d_{r_A r_i} = \sqrt{\left(x_{r_A} - x_{r_i}\right)^2 + \left(y_{r_A} - y_{r_i}\right)^2}$. Then, the average residual energy of $|Q|$ neighbor nodes is determined by

$$\bar{E}_{r\text{-res}} = \frac{\sum_{i=1}^{|Q|} E_{r\text{-res}_i}}{|Q|}, \tag{17.12}$$

where $E_{r\text{-res}_i} = E_{\text{init}} - \sum_{\text{round}=1}^{\infty} E_{r_i}$.

Due to the demand of minimizing transmission energy and retaining balance of the IIoT system, at the first stage of selecting the next-hop forwarder, the routing scheme utilizes the transmitting data distance from r_A to its neighbors and each neighbor's residual energy. If a neighbor satisfies the condition that its transmission distance is less than $\overline{d_{r_A r}}$ and its residual energy is greater than $\overline{E_{r\text{-res}}}$, it can take part in the set defined by NH as follows:

$$\mathcal{NH} = \left\{ r_i | d_{r_A r_i} \leq \bar{d}_{r_A r} \wedge E_{r\text{-res}_i} \geq \bar{E}_{r-\text{res}}, \quad r_i \in Q, \quad Q \subset \mathcal{R} \right\} \tag{17.13}$$

Relied on the role of hop count presented in Eq. (17.5), to degrade power consumption and end-to-end delay in the IIoT, among routers in NH, the potential forwarder of r_A is denoted as follows:

$$r_{\text{NH}} = \left\{ \{ r_i | \arg\min_{\mathcal{NH}} \{\text{hop counts}\}, \quad r_i \in \mathcal{NH}, \quad \mathcal{NH} \subset \mathcal{R} \right\}. \tag{17.14}$$

Eventually, in the completion of determining the next potential hop, r_A replies a modified INI-ROU message to indicate which router gets permission to forward data. The next routers are responsible for conducting the routing process until r_A's destination succeeds in receiving all data packets from it.

17.5 Performance Evaluation

17.5.1 IEEE 802.15.4a CSMA/CA Scheme for IIoT

According to [28–31], by using 14/16 channels in the ISM band, IEEE 802.15.4a assigns channels to each cluster to avoid inter-cluster interference. In addition, the IEEE 802.15.4a also provides Chirp Spread Spectrum (CSS) PHY which are widely employed to solve the energy consumption problem in a large-scale network, as mentioned in [32]. Thanks to a power-management mechanism defined by IEEE 802.15.4a, the beacon-enable mode can potentially reduce power consumption.

In terms of channel accessing, although TDMA was used to satisfy the reliability requirements in industrial applications [33] and the IoT [34–37], there has been several strict limits on the duration of TDMA slots, i.e., which is fixed to 10 ms in WirelessHART, and the TDMA frame length would be extended if there is an

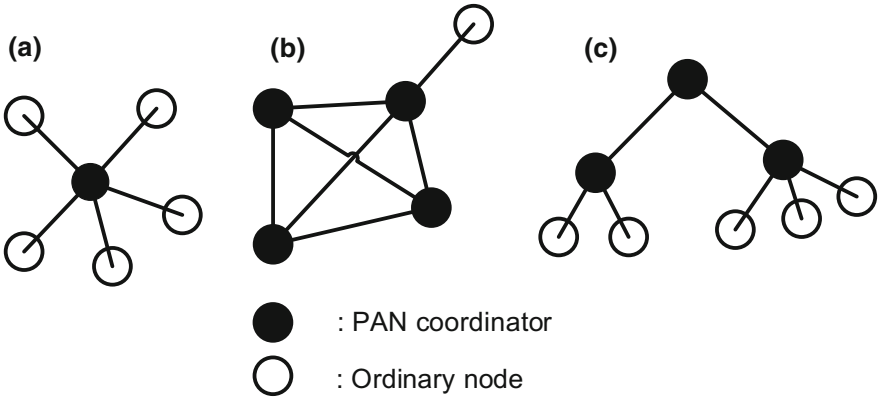


Fig. 17.8 Topology supported by IEEE 802.15.4: **a** star network, **b** mesh network, and **c** cluster tree

increase in the number of sensors in IIoT. Therefore, it is not as flexible as IEEE 802.15.4 in terms of superframe durations and beacon intervals. Moreover, based on Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) , the articles [38–40] demonstrated that IEEE 802.15.4 provides some advantages which make the real-time systems run accurately, compared to time division approaches. Thus, in this simulation, the IIoT elements operated IEEE 802.15.4a CSMA/CA protocol with beacon-enable mode as MAC layer while CSS was implemented at PHY layer.

In other words, IEEE standard 802.15.4 provides the fundamental lower network layer of a type of wireless personal area network (WPAN), which concentrates on low-cost, low-speed, ubiquitous communication between devices. As in [41], the basic components in an IEEE 802.15.4 network consist of a PAN coordinator and ordinary nodes. The PAN coordinator is responsible for managing the entire network, while ordinary nodes associate with the coordinator to participate in network operations. Regarding network topology, IEEE 802.15.4 not only supports the simple star network, but it also utilizes multi-hop topologies, such as cluster tree and mesh, as shown in Fig. 17.8. In this study, we deploy the proposed algorithms in a cluster-tree topology of the IEEE 802.15.4 standard to reveal the efficiency of the algorithms.

17.5.2 Simulation Model

ERRS is verified by OPNET Modeler. The IIoT system included 300 I/O devices, 100 routers, and 50 gateways. The gateway acted as PAN coordinators and the routers acted as coordinators. Meanwhile, all I/O devices acted as end devices associated with their coordinators. All IIoT elements were allocated randomly in a $500 \times 500\,m^2$ area. Table 17.3 details the main simulation parameters.

Table 17.3 Simulation parameters

Symbol	Definition
IIoT terrain size	500 m × 500 m
IIoT element distribution	Randomly
$\|\mathcal{S}\|$	300
$\|\mathcal{R}\|$	100
$\|\mathcal{G}\|$	50
$d_{s_{\max}}$, $d_{r_{\max}}$	[0; 15 m]
MAC layer	IEEE 802.15.4a CSMA/CA with *beacon-enable* mode
Duration of one backoff period slot for operations	320 μs
Beacon order ($0 \leq \text{BO} \leq 14$)	13
Beacon interval ($\text{BI} = 15.36 \times 2^{\text{BO}}$)	125.8 s
Superframe order ($0 \leq \text{SO} \leq \text{BO} \leq 14$)	7
Active period ($\text{SI} = 15.36 \times 2^{\text{SO}}$)	1.97 ms
PHY layer	CSS
Carrier-sensing range	30 m
Radio propagation	Two-way ground
Data rate	11 Mbps
Packet size	32 bytes
E_{elec}	50 nJ/bit
E_{DA}	5 nJ/bit/signal
E_{init}	2 J
ϵ_{fs}	10 pJ/bit/m^2
ϵ_{mp}	0.0013 pJ/bit/m^4
d_0	75 m

17.5.3 Simulation Results

To reveal the enhancement of ERRS, it was compared with Minimal Energy Consumption Algorithm (MECA) [24], Distributed Cluster Computing Energy-Efficient Routing Scheme (DCC) [42], and Energy-Aware Routing (ERA) [43] in terms of energy consumption, network lifetime, and end-to-end delay. ERA, ERRS is quite similar to, was limited to only Wireless Sensor Networks instead of larger scale and complex systems like the IIoT. Meanwhile, MECA and DCC were employed to qualify the power constraints in IoT systems. MECA utilized the optimization model to find the lowest energy path from the source to the destination, while DCC only focused on indicating how to form a cluster using residual energy. However,

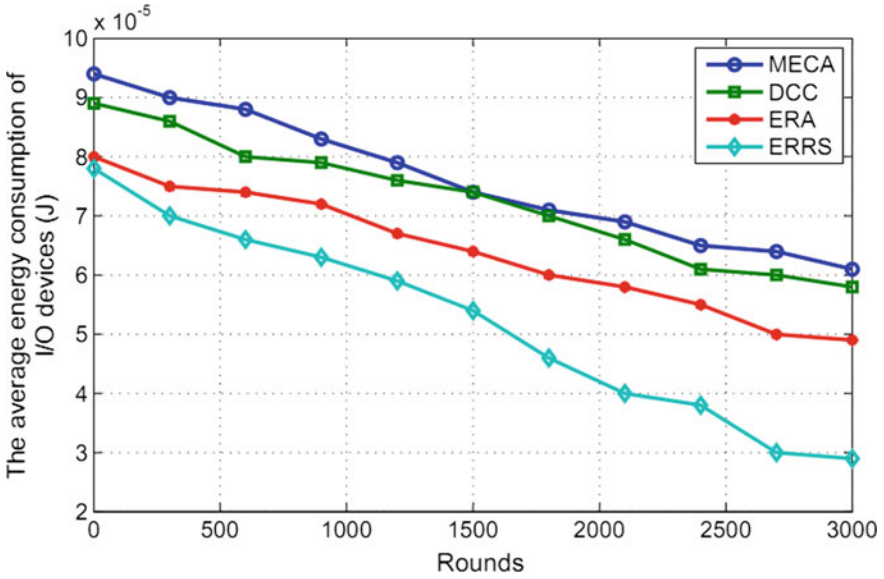


Fig. 17.9 Average energy consumption of I/O devices

the authors of MECA and DCC did not evaluate the impact of their schemes on end-to-end delay of their systems. Therefore, these previous works have several significant limitations. Thanks to an optimal routing implemented on IEEE 802.15.4a CSMA/CA, ERRS outperforms the conventional schemes in terms of power saving and end-to-end delay.

17.5.3.1 Energy Consumption

Figure 17.9 depicts the average energy consumed by I/O devices when increasing the number of rounds.

Of all four routing schemes reducing the energy consumption of I/O devices, ERRS can achieve the lowest energy dissipation, a gradual decrease from 7.8×10^{-5} to 2.9×10^{-5} J, in comparison with other methods.

MECA searched for a path that consumes the lowest energy to forward data using only Steiner tree and a simple K-means clustering method. Transmission distance was not considered as a metric for forming a cluster and a route. Consequently, the amount of power spent sending data by sensing nodes in MECA is highest among four routing schemes. The authors of DCC and ERA presented optimal methods that can create potential clusters using residual energy and distance. However, DCC deployed the algorithm in a topology that consisted of a BS while the scheme in ERA was proposed in WSNs. Therefore, when employed in large-scale IIoT systems,

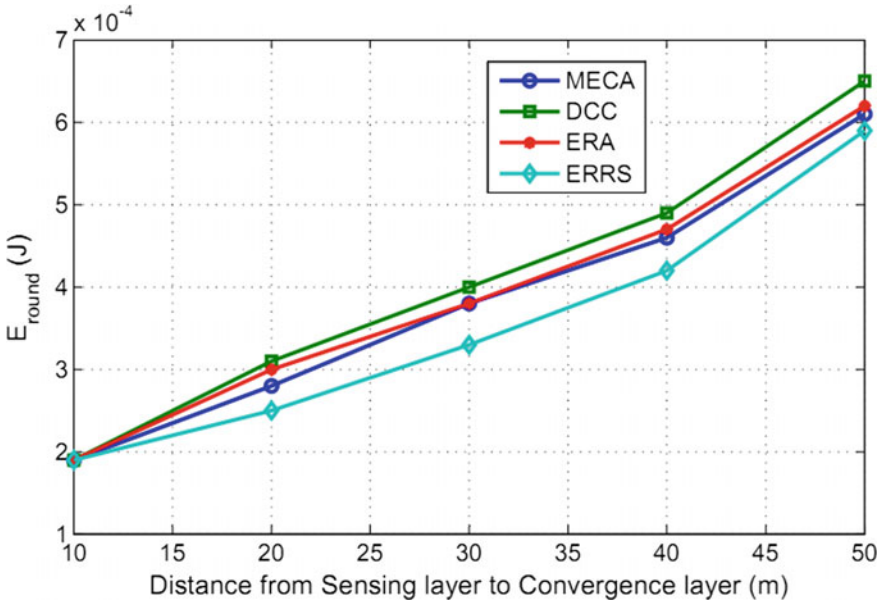


Fig. 17.10 E_{round} versus distance between sensing layer to convergence layer

obviously, each I/O device in the above algorithms cannot save power as much power as ERRS.

To further demonstrate the more positive impacts of ERRS, this study considered the effects of the four schemes into account on E_{round} when varying the distance between the sensing layer and the convergence layer. Owing to the large scale of IIoT systems, transmission distance has been a challenge for routing. Due to forwarding data directly to the destination instead of multi-hop paths, all the CHs in DCC need consumption of more transmission power than ERA and the application of a Steiner tree algorithm to search a path as shown in Fig. 17.10.

Owing to the advantages of IEEE 802.15.4a which are low power and long-communication range, ERRS still maintains the lowest energy consumption, which is in the range from 1.9×10^{-4} to 5.9×10^{-4} for a round when the distance from an I/O device to the gateway is extended.

17.5.3.2 Network Lifetime

The network lifetime of the four routing algorithms is illustrated clearly in Fig. 17.11. In this chapter, the lifetime of the network is referred to the number of alive nodes remaining operational after rounds of data transmission. As introduced in the system model, the first purpose of ERRS was reducing the data transmission power of I/O devices in a cluster. Thus, transmission distance and residual energy were considered

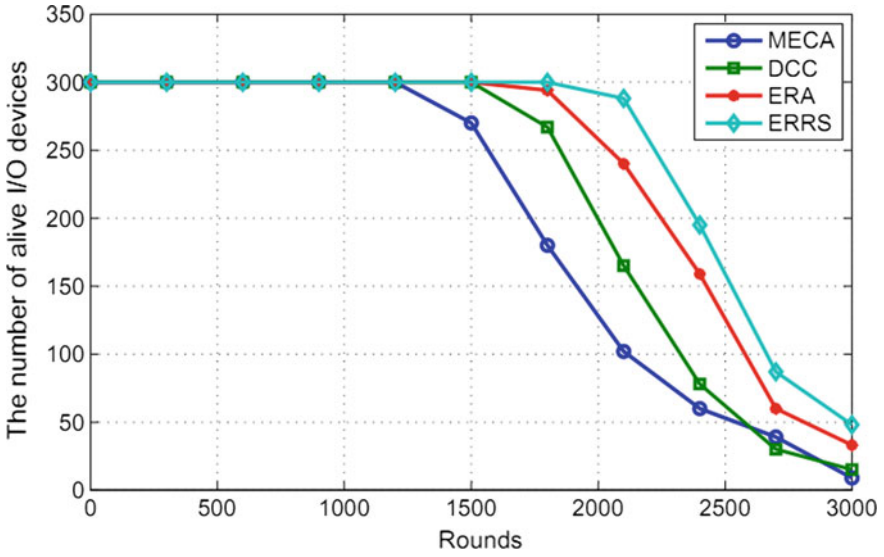


Fig. 17.11 The network lifespan

to create an optimal cluster structure. In addition, structuring the system into three-tiered layers characterized by different roles of different devices in layers, the traffic load of the network is decreased and balanced significantly, that results in power saving for all the devices in the systems.

Regarding selecting energy-efficient routing paths, ERRS not only selects the most reliable sink to complete the route, but was responsible for selecting the multi-router path that had the lowest energy dissipation to forward data from the I/O devices to the gateways. Moreover, because IEEE 802.15.4a could avoid negative effects of interference and collisions, ERRS outperformed all related in extending network lifetime.

17.5.3.3 Average End-to-End Delay

In the two algorithms described by MECA and DCC, the hop count factor was not taken into account as a condition for constructing a route. Meanwhile, to guarantee latency requirements for large-scale IIoT systems, ERRS in this study selected the potential forwarder with the lowest number of hop counts to a gateway. In large-scale complex networks like the IIoT, besides addressing the power-saving issue, demand for maintaining the lowest end-to end delay has been increasing. As a result, ERRS utilized the shortest path with the minimum hop count to decrease end to-end delay from I/O devices to gateways. In addition, when forwarding data simultaneously between clusters, unsuccessful transmission caused by collisions or inter-cluster interference were mitigated by IEEE 802.15.4a CSMA/CA algorithm.

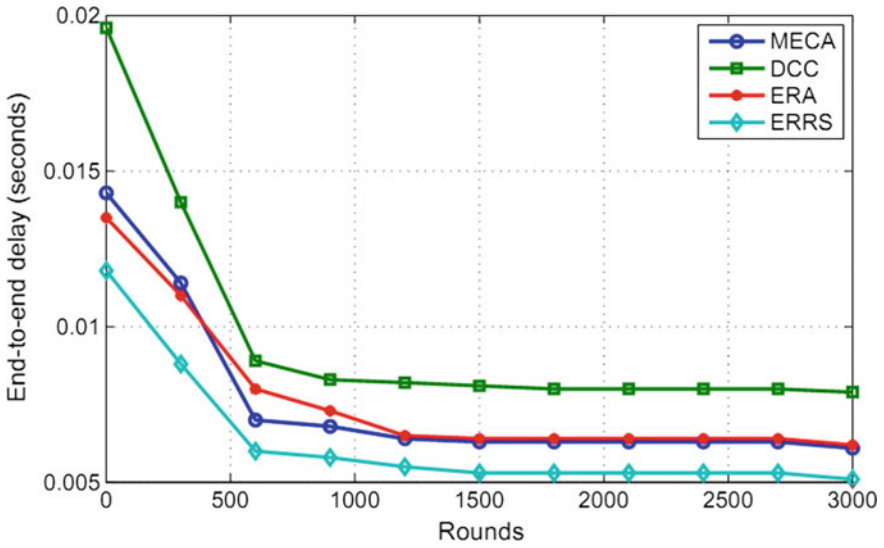


Fig. 17.12 Average end-to-end delay versus simulation time

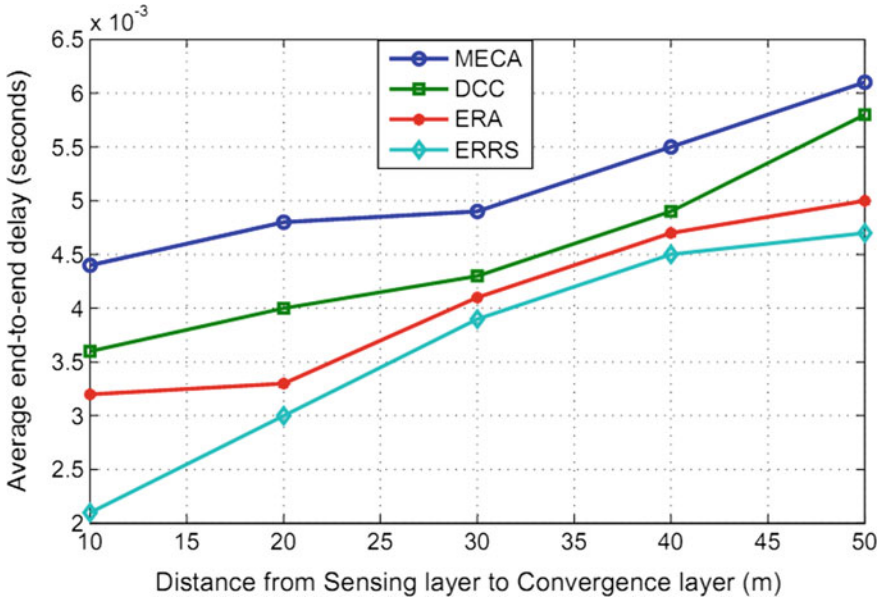


Fig. 17.13 Average end-to-end delay versus distance from sensing layer to convergence layer

As shown in Fig. 17.12, the end-to-end delay in ERRS remains steady at the lowest level ranging from 0.0118 s to 0.0051 s when compared to the other schemes.

Even when the distance between the sensing layer and the convergence layer was extended as in Fig. 17.13, the proposed approach still achieved the best outcome. Although the average end-to-end delay of all four routing schemes grew exponentially, the increase in the delay of ERRS was significantly lower than the remains.

17.6 Conclusions

This chapter proposed a routing approach that can minimize the entire energy consumption of large-scale complex IIoT. Relying on two crucial keys of transmission distance and residual energy, the routing approach overcame the disadvantages of the previous works. As well as proposing an optimal routing, the study also defined an energy-based clustering method employing a new parameter, namely, the selecting factor. The selecting factor was used by I/O devices to join an eligible cluster. Besides evaluating the efficiency of ERRS, the research also satisfied the latency requirements of IIoT systems utilizing hop count as a condition of constructing a data route. The simulation results showed that ERRS outperformed the conventional schemes in terms of power savings, network lifetime, and average end-to-end delay. As a future research direction, the efficiency of ERRS will be verified by implementation of existing devices in industrial factories to address practical concerns. And, to extend the contribution of this chapter, we will investigate ERRS in military and agricultural applications.

References

1. Xu LD, He W, Li S (2014) Internet of things in industries: a survey. IEEE Trans Industr Inf 10(4):2233–2243
2. Chao H, Chen Y, Wu J (2011) Power saving for machine to machine communications in cellular networks. In: 2011 IEEE GLOBECOM workshops (GC Wkshps), Dec 2011, pp 389–393
3. Dhondge K, Shorey R, Tew J (2016) HOLA: heuristic and opportunistic link selection algorithm for energy efficiency in industrial internet of things (IIoT) systems. In: 2016 8th international conference on communication systems and networks (COMSNETS), Jan 2016, pp 1–6
4. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
5. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun Surv Tutor 17(4):2347–2376
6. Li S, Xu LD, Zhao S (2015) The internet of things: a survey. Inf Syst Front 17(2):243–259
7. Wang K, Wang Y, Sun Y, Guo S, Wu J (2016) Green industrial internet of things architecture: an energy-efficient perspective. IEEE Commun Mag 54(12):48–54
8. Quang PTA, Kim D-S (2015) Clustering algorithm of hierarchical structures in large-scale wireless sensor and actuator networks. J Commun Netw 17(5):473–481
9. Nessa A, Kadoch M, Hu R, Rong B (2012) Towards reliable cooperative communications in clustered ad hoc networks. In: Global communications conference (GLOBECOM), 2012 IEEE, Dec 2012, pp 4090–4095

10. Vakil S, Dong M, Liang B (2010) Effect of cluster size selection on the throughput of multi-hop cooperative relay. In: Vehicular technology conference fall (VTC 2010-Fall), 2010 IEEE 72nd, Sept 2010, pp 1–5
11. Liu L, Hua C, Chen C, Guan X (2015) Relay selection for three-stage relaying scheme in clustered wireless networks. IEEE Trans Veh Technol 64(6):2398–2408
12. Palattella MR, Accettura N, Vilajosana X, Watteyne T, Grieco LA, Boggia G, Dohler M (2013) Standardized protocol stack for the internet of (important) things. IEEE Commun Surv Tutorials 15(3):1389–1406. Third 2013
13. Agha KA, Bertin MH, Dang T, Guitton A, Minet P, Val T, Viollet JB (2009) Which wireless technology for industrial wireless sensor networks? The development of OCARI technology. IEEE Trans Industr Electron 56(10):4266–4278
14. Silva FA (2014) Industrial wireless sensor networks: applications, protocols, and standards [book news]. IEEE Ind Electron Mag 8(4):67–68
15. Civerchia F, Bocchino S, Salvadori C, Rossi E, Maggiani L, Petracca M (2017) Industrial internet of things monitoring solution for advanced predictive maintenance applications. J Ind Inf Integr
16. Kruger CP, Hancke GP (2014) Implementing the internet of things vision in industrial wireless sensor networks. In: 12th IEEE international conference on industrial informatics (INDIN) 2014, July 2014, pp 627–632
17. Zhang D, Zhu Y, Zhao C, Dai W (2012) A new constructing approach for a weighted topology of wireless sensor networks based on local-world theory for the internet of things (IOT). Comput Math Appl 64(5):1044–1055 (Advanced Technologies in Computer, Consumer and Control)
18. Zhang D, Wang X, Song X, Zhang T, Zhu Y (2015) A new clustering routing method based on PECE for WSN. EURASIP J Wirel Commun Netw 2015(1):162
19. Zhang D, Zheng K, Zhang T, Wang X (2015) A novel multicast routing method with minimum transmission for WSN of cloud computing service. Soft Comput 19(7):1817–1827
20. Zhang D, Li G, Zheng K, Ming X, Pan ZH (2014) An energy balanced routing method based on forward-aware factor for wireless sensor networks. IEEE Trans Industr Inf 10(1):766–773
21. Laha A, Cao X, Shen W, Tian X, Cheng Y (2015) An energy efficient routing protocol for device-to-device based multihop smartphone networks. In: 2015 IEEE international conference on communications (ICC), June 2015, pp 5448–5453
22. Zhang Y, He S, Chen J (2016) Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks. IEEE/ACM Trans Netw 24(3):1632–1646
23. Ben Arbia D, Alam MM, Attia R, Ben Hamida E (2017) ORACE-Net: a novel multi-hop body-to-body routing protocol for public safety networks. Peer-to-Peer Netw Appl 10(3):726–749
24. Huang J, Meng Y, Gong X, Liu Y, Duan Q (2014) A novel deployment scheme for green internet of things. IEEE Internet Things J 1(2):196–205
25. Ben Arbia D, Alam M, Kadri A, Ben Hamida E, Attia R (2017) Enhanced IoT-based end-to-end emergency and disaster relief system. J Sens Actuator Netw 6(3):19
26. Heinzelman W, Chandrakasan A, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wirel Commun 1(4):660–670
27. Long NB, Nhon T, Kim DS (2016) Rate-estimation-based relay selection scheme for large-scale wireless networks. IET Commun 10(12):1501–1507
28. Son ND, Tan DD, Kim D-S (2012) Backoff algorithm for time critical sporadic data in industrial wireless sensor networks. In: International conference on advanced technologies for communications (ATC), 2012, Oct 2012, pp 255–258
29. Tavakoli H, Miic J, Naderi M, Miic V (2013) Energy-efficient clustering in IEEE 802.15.4 wireless sensor networks. In: 33rd IEEE international conference on distributed computing systems workshops (ICDCSW), July 2013, pp 262–267
30. Anastasi G, Conti M, Di Francesco M (2011) A comprehensive analysis of the MAC unreliability problem in IEEE 802.15.4 wireless sensor networks. IEEE Trans Industr Inf 7(1):52–65
31. Quang PTA, Kim D-S (2014) Throughput-aware routing for industrial sensor networks: application to ISA100.11a. IEEE Trans Industr Inf 10(1):351–363

32. Karapistoli E, Pavlidou F-N, Gragopoulos I, Tsetsinas I (2010) An overview of the IEEE 802.15.4a standard. IEEE Commun Mag 48(1):47–53
33. Chen D, Nixon M, Mok A (2010) WirelessHART: real-time mesh network for industrial automation, 1st edn. Springer Publishing Company, Incorporated
34. Yang D, Gidlund M, Shen W, Xu Y, Zhang T, Zhang H (2013) CCA embedded TDMA enabling acyclic traffic in industrial wireless sensor networks. Ad Hoc Netw 11(3):1105–1121
35. Yang Y, Cao S (2014) Multiplex TDMA link assignment with varying number of sensors in industrial wireless sensor networks. In: 2014 international conference on identification, information and knowledge in the internet of things, Oct 2014, pp 242–247
36. Zhai C, Zou Z, Chen Q, Xu L, Zheng L-R, Tenhunen H (2016) Delay-aware and reliability-aware contention-free MF-TDMA protocol for automated RFID monitoring in industrial IoT. J Ind Inf Integr 3:8–19
37. Pielli C, Biason A, Zanella A, Zorzi M (2016) Joint optimization of energy efficiency and data compression in TDMA-based medium access control for the IoT. In: 2016 IEEE GLOBECOM workshops (GC Wkshps), Dec 2016, pp 1–6
38. Yoo S, Chong PK, Kim D, Doh Y, Pham ML, Choi E, Huh J (2010) Guaranteeing real-time services for industrial wireless sensor networks with IEEE 802.15.4. IEEE Trans Ind Electron 57(11):3868–3876
39. Tang C, Song L, Balasubramani J, Wu S, Biaz S, Yang Q, Wang H (2014) Comparative investigation on CSMA/CA-based opportunistic random access for internet of things. IEEE Internet Things J 1(2):171–179
40. Du W, Navarro D, Mieyeville F (2015) Performance evaluation of IEEE 802.15.4 sensor networks in industrial applications. Int J Commun Syst 28(10):1657–1674
41. Anastasi G, Conti M, Francesco MD, Neri V (2010) Reliability and energy efficiency in multi-hop IEEE 802.15.4/ZigBee wireless sensor networks. In: 2010 IEEE symposium on computers and communications (ISCC), June 2010, pp 336–341
42. Chang J-Y (2015) A distributed cluster computing energy-efficient routing scheme for internet of things systems. Wirel Pers Commun 82(2):757–776
43. Amgoth T, Jana PK (2015) Energy-aware routing algorithm for wireless sensor networks. Comput Electr Eng 41:357–367

Chapter 18
3D Perception Framework for Stacked Container Layout in the Physical Internet

18.1 Introduction

The way goods are moved has become more and more inefficient with a huge impact on the environment. In the work [10], 13 key symptoms are reported to highlight the unsustainability of the current logistics systems. Toward the future visions, the work claims that such traditional logistics paradigm is no longer working and cannot meet the requirements of modern society. That is why it has increasing amounts of green initiatives at tactical, strategic, and operational levels of logistics [8]. However, these initiatives often seem limited to specific applications for a single firm or within a limited number of chain partners, rather than across the entire supply chain. In this context, a new integrated logistics paradigm is necessary by correlating economic, ecological, and social aspects, and also including all stakeholders of the supply chain. The Physical Internet is becoming an extremely promising solution to the global logistics sustainability grand challenge. This initiative proposes to design a system to move, handle, store, realize, supply, and use physical objects throughout the world in a manner that improves efficiency, effectiveness, and sustainability simultaneously [2]. This paradigm shift is mainly founded on the interconnection of logistic networks and the encapsulation concept, in a similar way to the internetworking and transport of packets in the Digital Internet. The latter encapsulates information in standardized data packets. Besides, all interfaces and protocols are designed and developed independently so as to exploit this encapsulation properly. In this way, data packets can be processed by different network equipment (e.g., routers or switches), and carried through different networks using various types of media. By analogy, the Physical Internet does not manipulate physical goods directly, but standardized modular containers (called π-containers) that encapsulate physical merchandises, and composite π-containers which can be composed of a set of unitary and smaller π-containers. The interfaces and protocols of Physical Internet are designed through π-facilities (such as π-hubs, π-movers, π-carriers, etc.) to obtain an efficient and sustainable universal interconnectivity [16]. Therefore, the core of

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_18

the Physical Internet concept is the handling of π-containers throughout an open global logistic infrastructure.

This vision will require rethinking the global supply chain, where modular π-containers will be manipulated over time (transport, store, load/unload, build/dismantle …) but also, their subparts may be changed among the different nodes of the Physical Internet network (e.g., partial loading/unloading, containers splitting and merging). This composition/decomposition of π-containers is a key element to generate an efficient and sustainable global logistics chain [9]. Moreover, the Physical Internet is expected to exploit the capabilities of smart containers as much as possible, in order to enable decision-making processes on the spot that will open new opportunities such as real-time routing [15]. For instance, π-containers could adapt their routing plans in each π-hub, given new current information on opportunities and constraints. As a result, a significant challenge is then to maintain the π-container traceability in a highly dynamic transport and logistic system.

Although the constitution of a composite π-container can be assumed to be perfectly known when it was set up, the multiplicity of various routing and transformation processes in Physical Internet can introduce a desynchronization between the physical and informational flows. To overcome this problem, we propose a system able to generate and maintain automatically a virtual 3D layout reflecting the spatial distribution of π-containers. The objective is to identify but above all to be able to locate the exact position of stacked items. To do this, the reliance on wireless technologies and localization techniques is needed. However, the originality of our approach is to be independent of the quality of received signals, which is important in harsh environments and operating conditions encountered in logistics. Besides, no pre-installed localization infrastructure is required. Our approach is based on the use of smart π-container embedding wireless sensor nodes and the knowledge extraction from the obtained spontaneous network. Once information collected, a Constraint Satisfaction Problem (CSP) can be formulated, where each feasible solution of the CSP is a potential loading pattern. The resulting composite container model can be used to provide up-to-date information (permanent inventory) but also to detect any error during the encapsulation process, which might have a negative impact on the overall effectiveness and efficiency of the Physical Internet.

The remainder of this chapter is organized as follows: Sect. 18.2 summarizes the previous works related to localization technologies and techniques commonly used in logistics management activities. Section 18.3 formally defines the problem and the methodology to generate the virtual 3D layout of a composite π-container, followed by the mathematical formulation of the CSP problem in Sect. 18.4. As a proof-of-concept, application and results are given in Sect. 18.5, where generated problem instances are utilized. Finally, some conclusions and perspectives are provided in Sect. 18.6.

18.2 Literature Review

Inbound and outbound logistics systems typically include activities such as transportation and distribution, inventory, warehousing, and materials handling. To improve performance and increase productivity, the use of automation in supply chain processes has expanded dramatically in recent years. For instance, maximum productivity can be obtained in material handling through minimizing the number and complexity of handling operations. To do this, it is desirable for all the movements to be as simple and automated as possible [7]. Similarly, the use of ICT (Information and Communication Technology) provides more information availability and visibility in the supply chain management for transaction execution, collaboration/coordination and decision [17]. For that purpose, the reliance on identification and localization technologies into logistics processes is prominent.

Technologies and devices, such as bar code readers, Radio-Frequency Identification (RFID) tags, have revolutionized the way to automatically and continuously identify logistics objects. These automatic identification systems also known as Auto-ID technologies are mainly used to detect the presence of nearby objects. In this process, tagged objects equipped with a barcode label or a RFID chip are localized when they are in proximity to optical or RF readers, respectively. Although Auto-ID technologies have been proven to be sufficiently adequate for localization in some logistics applications, such as inventory management, there are numerous other applications that cannot benefit from this technology due to some limitations. These limitations are mainly related to the environment in which tags and readers communicate [13], and also the location accuracy reduced to the presence or not of the object in the range. The localization precision also depends on the type of tags (passive, active, and semi-passive tags) and the number of readers used.

To overcome the problem concerning the lack of accuracy, several approaches have been proposed in the literature. Most of them rely on the combination of Auto-ID systems with other positioning technologies and techniques. This is certainly due to the large deployment of Auto-ID solutions in all logistics activities. For instance, the authors in [5] propose a hybrid RFID-GPS system to realize continuous monitoring and precise localization in a vehicle terminal. The data collected by the RFID reader are combined with GPS position to track and trace mobile equipment in real time. When GPS signals are not available like in an indoor environment, other signals can be used. A highly accurate indoor location system (SmartLocator) using infrared signals and RFID is introduced in [4]. The solution increases the efficiency of warehouse input/output operation and enables the effective utilization of storage spaces in a completely free (or random space allocation) warehouse. Thiesse, Fleisch, and Dierkes in their work [19] proposed a real-time localization system (LotTrack) which combines RFID and ultrasound sensors. Their approach based on trilateration method (like GPS solutions) is used to improve tracking visibility for inbound logistics in a chip-manufacturing company. In the previous works location estimation techniques are used to determine the position, in particular, trilateration and proximity. Recently distance measurement conditions, the authors have obtained, with

standard UHF passive RFID, acceptable accuracy which can satisfy the localization requirement in warehouse environments [20].

Other localization technologies have also been experimented such as Wireless Sensor Networks. WSNs offer a number of advantages over RFID implementations such as multi-hop communication, sensing capabilities, and programmable sensor nodes. Moreover, the growing trend of smart objects in logistics enables the development of these new localization solutions. For instance, the localization of euro-pallets within a high bay warehouse is proposed in [18]. Sensor devices based on the wireless 802.15.4 technology are installed on pallets and four anchors are fixed to the racks. The distance measurement is obtained using signal propagation delay and an Extended Kalman Filter to reduce the noise and obtain a precise location. The authors employed the same technique to track forklift trucks in the warehouse [14]. A WSN approach has been studied with outdoor localization of shipping containers in ports and terminals [1]. Each container is equipped with six nodes and uses RSS measurements combined with the geometrical constraints to determine the relative position of containers.

From the above-related works, we propose in this chapter a nonconventional WSN approach where no distance measurement is needed. We exploit the neighborhood information between nodes of the Wireless Sensor Network. This information is used to determine the physical layout of a composite π-container in the Physical Internet.

18.3 Problem and Methodology

18.3.1 Problem Definition and Proposed Approach

As mentioned in the introduction, the Physical Internet manipulates modular π-containers through an open global logistic infrastructure, including key facilities such as π-hubs. The π-containers are organized into three functional categories: transport, handling, and packaging containers, with their modular dimensions [11]. Although the set of modular dimensions must be subject to an international standard committee so as to obtain maximum consensus, the partners of the CELDi PI project [6] developed a mathematical model to determine the best container size to maximize space utilization. They demonstrated that a set of external dimensions including the following values in meters {0.12, 0.24, 0.36, 0.48, 0.6, 1.2, 2.4, 3.6, 4.8, 6, 12}, increases the space utilization at the unit load level. These π-containers can be composed and interlocked to build "composite" π-containers and later decomposed into sets of unitary π-containers, as depicted in Fig. 18.1. The π-containers are key elements of the Physical Internet and we propose in this chapter, a π-container traceability system for a better synchronization of the physical and informational flows during the composition/decomposition processes. Such a system can provide information availability and visibility, and thus contributes to the efficiency of a future open global logistics system.

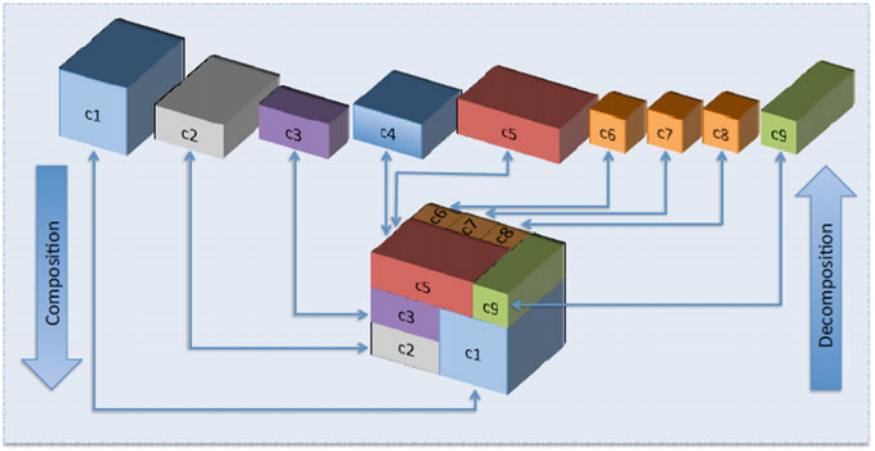


Fig. 18.1 Example of composition/decomposition of PI-containers sourced from [12]

The objective of this work is therefore to generate a virtual 3D layout reflecting the physical layout of a composite π-container (with the exact position of stacked unitary π-containers). For that purpose, we consider that each π-container is a smart object according to the key functional specifications of π-containers [10]. In other words, a π-container is equipped with low-cost sensors and communication devices to be able to sense and measure its environment, and communicate with other smart containers. An example of the integration of all these capabilities is the Intelligent Container (so-called InBin) recently developed by the Fraunhofer Institute for Material Flow and Logistics. We assume in our approach that a wireless sensor node is attached to one corner of each container and stores information about the container such as the container category, the identifier, and its dimensions. In a consistent manner, one sensor node embedded at the composite container level, acts as a gateway and provides the interface between the management information system and the composite container. This node will be also used as an absolute reference to determinate the virtual 3D layout. According to the transmission range of the sensor nodes, the composition of a composite π-container will establish a spontaneous wireless network, as shown on the left side of Fig. 18.2. As a result, each π-container can collect information about its neighborhood and compute its neighbor list. Once collected at the gateway, a Constraint Satisfaction Problem (CSP) can be formulated where:

- The neighbor table gives position constraints between the unitary π-containers, but also with the composite π-container. It represents allocation restrictions in the CSP problem.
- The container size provides basic geometric constraints. The unitary π-containers lie entirely within the composite container and do not overlap. Each one may only be placed with their edges parallel to the walls of the composite π-container.

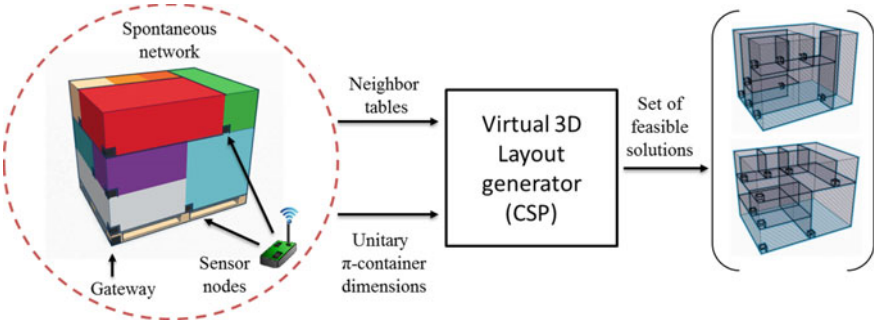


Fig. 18.2 General overview of the proposed approach

Also, each feasible solution of the CSP is a potential loading pattern, and the resulting virtual 3D layout can be generated through a graphical user interface. The Figure illustrates the general approach proposed in this chapter.

18.3.2 Methodology and Assumptions

The approach proposed above relies on two main principles. The first concerns the wireless sensor nodes which are programmable and can be synchronized to execute specific tasks that include collaboration among local neighbors. The gateway can play a role of coordinator in the network. The second one is related to the transmission range of the nodes which is influenced by several factors like the operating frequency, the transmission power or the antenna gain. We consider here the transmission range as the distance where the communication between two nodes can be set up, and beyond which the quality of the communication exceeds a given stability. We assume that each node uses one omnidirectional antenna and can change its transmission range through the control of the transmission power level. To verify that this assumption is reasonable, we realized some experiments using a MICAZ platform. The nodes are based on a 2.4 GHz IEEE 802.15.4 compliant RF transceiver offering 32 transmission power levels. The experimentation has been realized under different conditions (e.g., sensor's line-of-sight or obstacles). A communication distance from a few centimeters to several dozen meters is achieved in consonance with the range of the RF power values. The detailed results being out of the scope of the chapter, they are not given here. To conclude, each unitary π-container (with the external dimensions previously given in Sect. 18.3.1) can adapt its transmission range and detect the proximity of other π-containers in its environment, and communicate with them.

As a result, position constraints between the unitary π-containers which will be used in our CSP, can be obtained after several successive operations performed in collaboration at the gateway and node levels:

(1) The first one concerns the identification of the number of unitary π-container that composed the composite π-container. This can be achieved by the implementation of a discovery mechanism to count the number of nodes in the network. This task initiated by the gateway is performed by broadcasting an announcement request, and after receiving it, each node replies with a packet that contains the identifier of the π-container. The identifier can also be used to ensure that the detected π-container belongs to the composite π-container. If the transmission range is smaller than the dimension of the composite π-container, all containers cannot be discovered at the same time. On the other hand, the protocol must be adapted to a multi-hop algorithm needed to discover the entire network. The algorithm can be repeated at a different time to be tolerant to possible packet losses.
(2) After that, each node detects and identifies the proximity of other nodes. The same discovery algorithm can be used to search neighbors. A node broadcasts discovery packets so that they reach all nodes in the neighborhood. At this step, we only assume bidirectional communication links. That is, sensor nodes i and j are neighbors if i can receive j's messages and vice versa. The algorithm can also be repeated at a different time to be tolerant to possible interferences. The symmetric neighbor relation implies that sensor nodes use the same transmission power level. Each node can compute a local neighbour list that gives the relative position between the unitary π-containers. This information is essential to generate pertinent allocation constraints for our CSP, thereby reducing the number of feasible solutions. This point will be discussed in the following of this section.
(3) The node attached to the composite π-container can be used as a reference of our localization system. Hence, the position of all the π-container can be derived from this absolute position. Forwarding mechanisms based on a multi-hop routing tree are commonly used in Wireless Sensor Networks to transfer collected data throughout the network. Such a technique is used to collect the local neighbor lists at the gateway where they are combined together to generate a global neighbor list, before to be expressed as position constraints in our CSP.

The advantage of our approach is that the properties of the received signals, e.g., signal strength or angle of arrival, are not considered in the successive operations to localize the π-container. The position is only determined from geometrical and neighborhood information of the π-containers, which is important in harsh environments and operating conditions encountered in logistics. However, the number of feasible solutions issued from the CSP solver is directly linked to the pertinence of the global neighbor list, and therefore, the choice of the transmission range is important.

Indeed, two nodes are neighbors if they can communicate, i.e., the distance between them is lesser than or equal to the transmission range. Therefore, with a transmission range smaller than the smallest dimension of the π-containers, a lot of nodes will be unable to find neighbors in close proximity. This is illustrated in Fig. 18.3 where a composite π-container is made up of five π-containers (πc_1, πc_2

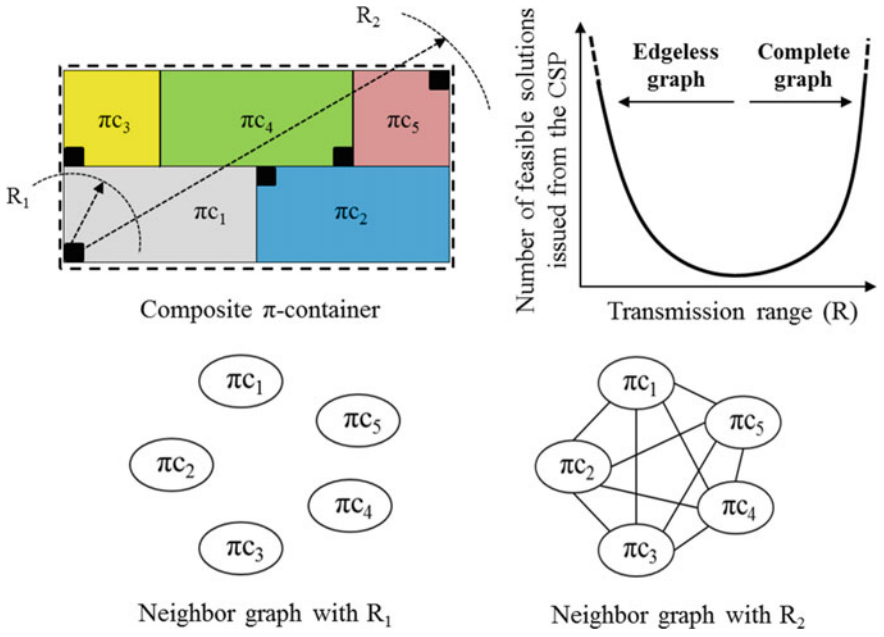


Fig. 18.3 Influence of the transmission range on the number of solutions

… πc_5). The distribution of the sensor nodes (black squares) corresponds to the arrangement obtained when the composite π-container is set up. The global neighbor list is modeled by an undirected graph for different transmission range R_i. A vertex represents a sensor node belonging to a π-container, and edges, the neighbor relationships between them. The neighbor graph obtained with a transmission range R_1, inferior to the smallest dimension of the π-containers, is an edgeless graph (with no edges). Therefore, it will be impossible to specify position constraints in our CSP. The number of feasible solutions, i.e., virtual 3D layouts for the composite π-container, will correspond to every possible arrangement of the π-containers into the composite π-container. Similarly, if the transmission range is superior to the length of the composite π-container diagonal, all nodes are neighbors one of another. The neighbor graph will be a complete graph, as depicted in Fig. 18.3 with R_2. A lot of solutions, where the position constraints will be satisfied, can be found from a simple permutation or rotation of π-containers. Thus, the transmission range must be adapted according to the π-containers dimensions in order to obtain pertinent and sufficient position constraints.

18.4 Mathematical Formulation of the CSP Problem

Assume that a set of π-containers with modular dimensions is currently stacked in a composite π-container with given external dimensions. A Cartesian coordinate system is fixed to the Front-Left Bottom corner (FLB) of the composite π-container and defines the origin and orientation of the three axes of a three-dimensional space. We also designate the FLB corner of each π-container as the origin of its own coordinate system. This position denotes the orientation of a π-container, and corresponds to the initial location of the wireless sensor node. A set of neighborhood relations is defined from the global neighbor graph for a given value of the transmission range. In addition, we consider that π-containers can be freely rotated into different orientations. The objective of the CSP problem is to find the absolute coordinates and the orientation of each π-container, while satisfying the following set of constraints:

- Each π-container is stacked in the composite π-container.
- Each π-container must be orthogonally placed into the container.
- The overlap among the π-containers is prohibited.
- All neighborhood relations about the unitary π-containers and the composite π-container must be respected.

A solution to the CSP is a complete assignment that satisfies all the constraints. The following sections provide a summary of parameters and variables, and include a mathematical model of the Constraint Satisfaction Problem. This model is based on the mathematical formulation proposed by Chen et al. [3] to formulate the general container loading problem. This model has been adapted to consider the neighborhood constraints.

18.4.1 Parameters and Variables

18.4.1.1 Parameters

n	Total number of unitary π-containers stacked in the composite π-container (determined from the counting algorithm given in Sect. 18.3.2).
πc_i	π-container set, where $i \in \{0, \ldots, n\}$ and πc_0 is the composite π-container.
(x_o, y_o, z_o)	Origin of the Cartesian coordinate system, with $x_o = y_o = z_o = 0$.
(L_i, W_i, H_i)	Nonnegative external dimensions indicating the length, width, and height of the πc_i.
(L_c, W_c, H_c)	Nonnegative external dimensions indicating the length, width, and height of the composite π-container.
R	Transmission range of sensor nodes.
V	$i \times j$ matrix with i and $j \in \{0, \ldots, n\}$

The matrix elements V_{ij} denote the neighbor relationship between containers πc_i and πc_j. V_{ij} is equal to 0 if $i=j$. The remaining elements of the matrix V are deduced from the global neighbor list with: V_{ij} is equal to 1 if πc_i and πc_j are neighbors; otherwise, it is equal to -1.

18.4.1.2 Variables

(x_i, y_i, z_i)	Continuous variables indicating the coordinates of the FLB corner of container πc_i.
(x_{ni}, y_{ni}, z_{ni})	Continuous variables indicating the coordinates of the sensor node fixed to container πc_i.
(l_{xi}, l_{yi}, l_{zi})	Binary variables indicating whether the length of container πc_i is parallel to the *X-axis*, *Y-axis*, or *Z-axis*. The value of l_{xi} is equal to 1 if the length of container πc_i is parallel to the *X-axis*; otherwise, it is equal to 0.
(w_{xi}, w_{yi}, w_{zi})	Binary variables indicating whether the width of container πc_i is parallel to the *X-axis*, *Y-axis*, or *Z-axis*. The value of w_{xi} is equal to 1 if the width of container πc_i is parallel to the *X-axis*; otherwise, it is equal to 0.
(h_{xi}, h_{yi}, h_{zi})	Binary variables indicating whether the height of container πc_i is parallel to the *X-axis*, *Y-axis*, or *Z-axis*. The value of h_{xi} is equal to 1 if the height of container πc_i is parallel to the *X-axis*; otherwise, it is equal to 0.

It is clear that the binary variables, l_{xi}, l_{yi}, l_{zi}, w_{xi}, w_{yi}, w_{zi}, h_{xi}, h_{yi}, and h_{zi}, are dependent and determine the orientation of container πc_i. For example, if the length (L_i), width (W_i), and height (H_i) of the πc_i are parallel to the X, Y, and Z axes, respectively, then l_{xi}, w_{yi} and h_{zi} are equal to 1 and all the other variables are null. Consequently, the relationships between these variables need to be verified.

For each pair of π-containers (πc_i, πc_j), there is a set of six binary variables (Right_{ij}, Left_{ij}, Behind_{ij}, Front_{ij}, Below_{ij}, and Above_{ij}) that defines their relative positions in the composite π-container. Binary variables will be 1 if the container πc_j is to the right of, to the left of, behind, in front of, below, or above the container πc_i, respectively; otherwise, 0. These variables will be used to ensure that containers do not overlap.

18.4.2 Formulation

From the parameters and variables described above, the constraints of the CSP problem are as follows:

$$\forall i, j, 1 \le i, j \le n : \text{Right}_{ij} \cdot [x_j - (x_i + l_{xi} \cdot L_i + w_{xi} \cdot W_i + h_{xi} \cdot H_i)] \ge 0 \tag{18.1}$$

$$\forall i, j, 1 \le i, j \le n : \text{Left}_{ij} \cdot [x_i - (x_j + l_{xj} \cdot L_j + w_{xj} \cdot W_j + h_{xj} \cdot H_j)] \ge 0 \tag{18.2}$$

$$\forall i, j, 1 \le i, j \le n : \text{Behind}_{ij} \cdot [y_j - (y_i + l_{yi} \cdot L_i + w_{yi} \cdot W_i + h_{yi} \cdot H_i)] \ge 0 \tag{18.3}$$

$$\forall i, j, 1 \le i, j \le n : \text{Front}_{ij} \cdot [y_i - (y_j + l_{yj} \cdot L_j + w_{yj} \cdot W_j + h_{yj} \cdot H_j)] \ge 0 \tag{18.4}$$

$$\forall i, j, 1 \le i, j \le n : \text{Below}_{ij} \cdot [z_j - (z_i + l_{zi} \cdot L_i + w_{zi} \cdot W_i + h_{zi} \cdot H_i)] \ge 0 \tag{18.5}$$

$$\forall i, j, 1 \le i, j \le n : \text{Above}_{ij} \cdot [z_i - (z_j + l_{zj} \cdot L_j + w_{zj} \cdot W_j + h_{zj} \cdot H_j)] \ge 0 \tag{18.6}$$

$$\forall i, j, 1 \le i, j \le n : \text{Right}_{ij} + \text{Left}_{ij} + \text{Behind}_{ij} + \text{Front}_{ij} + \text{Below}_{ij} + \text{Above}_{ij} \ge 1 \tag{18.7}$$

$$\forall i,\ 1 \le i \le n : x_i + l_{xi} \cdot L_i + w_{xi} \cdot W_i + h_{xi} \cdot H_i \le L_c \tag{18.8}$$

$$\forall i,\ 1 \le i \le n : y_i + l_{yi} \cdot L_i + w_{yi} \cdot W_i + h_{yi} \cdot H_i \le W_c \tag{18.9}$$

$$\forall i,\ 1 \le i \le n : z_i + l_{zi} \cdot L_i + w_{zi} \cdot W_i + h_{zi} \cdot H_i \le H_c \tag{18.10}$$

$$\forall i,\ 1 \le i \le n : ((x_i + l_{xi} \cdot L_i + w_{xi} \cdot W_i + h_{xi} \cdot H_i) - x_{ni}) \cdot (x_i - x_{ni}) = 0 \tag{18.11}$$

$$\forall i,\ 1 \le i \le n : ((y_i + l_{yi} \cdot L_i + w_{yi} \cdot W_i + h_{yi} \cdot H_i) - y_{ni}) \cdot (y_i - y_{ni}) = 0 \tag{18.12}$$

$$\forall i,\ 1 \le i \le n : ((z_i + l_{zi} \cdot L_i + w_{zi} \cdot W_i + h_{zi} \cdot H_i) - z_{ni}) \cdot (z_i - z_{ni}) = 0 \tag{18.13}$$

$$\forall i, j,\ 0 \le i, j \le n : V_{ij} \sqrt{((x_{ni} - x_{nj})^2 + (y_{ni} - y_{nj})^2 + (z_{ni} - z_{nj})^2)} \le V_{ij} R \tag{18.14}$$

$$\forall i,\ 1 \le i \le n : l_{xi} + l_{yi} + l_{zi} = 1 \tag{18.15}$$

$$\forall i,\ 1 \le i \le n : w_{xi} + w_{yi} + w_{zi} = 1 \tag{18.16}$$

$$\forall i,\ 1 \le i \le n : h_{xi} + h_{yi} + h_{zi} = 1 \tag{18.17}$$

$$\forall i,\ 1 \le i \le n : l_{xi} + w_{xi} + h_{xi} = 1 \tag{18.18}$$

$$\forall i, 1 \le i \le n : l_{yi} + w_{yi} + h_{yi} = 1 \tag{18.19}$$

$$\forall i,\ 1 \le i \le n : l_{zi} + w_{zi} + h_{zi} = 1. \tag{18.20}$$

We can distinguish the constraints related to the neighborhood relations between nodes and those related to geometric feasibility conditions to place the π-containers. The constraints (18.1)–(18.7) are nonoverlapping conditions and certify that π-

containers do not overlap. Constraints (18.8)–(18.10) ensure that all π-containers are within the composite π-container. Constraints (18.11)–(18.13) represent the relation between the coordinates of the FLB corner and of the sensor node. The sensor node can potentially be placed at one the eight corners of the π-container according to π-containers can be freely rotated. Constraint (18.14) certifies that all neighbor relationships have been respected. Finally, constraints (18.15)–(18.20) ensure that the binary variables which determine the position of the π-containers are properly controlled to reflect practical positions. This model leads to find the absolute coordinates and the orientation of each π-container, which satisfy the neighbor relationships between the sensor nodes.

18.5 Application and Results

18.5.1 Experimental Setup

To validate the above approach and model, we consider scenarios that represent real compositions of a composite π-container. The objective is to find the spatial distribution of staked π-containers. As a proof-of-concept, a set of nine modular π-containers are presented in [9]. As mentioned in Sect. 3.1, the external dimensions of π-containers represent a large scale of values in order to cover all the functions, namely transport, handling, and packaging. Moreover a transmission range from a few centimeters to several dozen meters can be currently achieved with wireless sensor platforms. That's why a set of normalized values {1, 2, 3, 4} for the π-container dimensions are used, as shown Fig. 18.4. The sensor nodes (black squares) are initially fixed to the front-left bottom corner of the π-containers, and the origin of our Cartesian coordinate system is defined by the node of the composite π-container. We defined different scenarios where all the π-containers are stacked so that they fit perfectly the volume of a composite π-container with the dimensions $[4 \times 3 \times 3]$. They differ by the amount, the heterogeneity and the final orientation of π-containers stacked in the composite π-container.

The scenarios are illustrated in Fig. 18.5. A strongly heterogeneous set of 9 π-containers is used in scenarios 1 and 2, unlike scenario 3, which considers 18 identical π-containers with $[2 \times 1 \times 1]$ dimensions. In scenarios 1 and 3, the FLB corner

Fig. 18.4 Dimensions of the nine unitary and composite π-containers

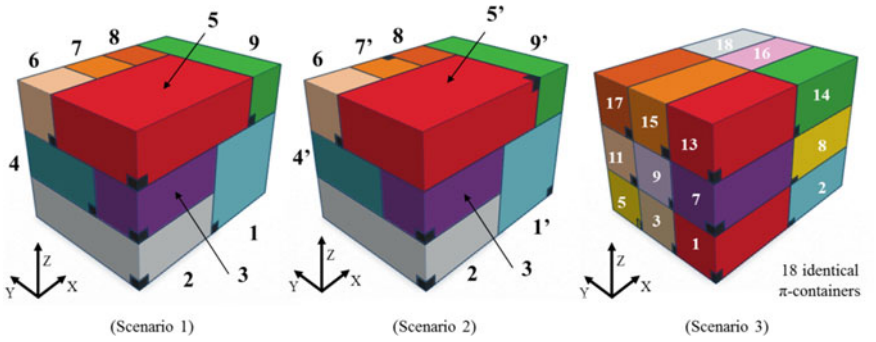


Fig. 18.5 Three composite π-containers used to validate the proposed approach

coordinates are identical to those of the sensor nodes. This can occur when composite π-containers were set up respecting orientation rules (e.g., with a "This way up!" sign) during the stacking process. No rotation of the π-containers is allowed in these scenarios. That means that, x_i, y_i, and z_i are equal to x_{ni}, y_{ni}, and z_{ni}, respectively. This constraint will reduce the research space and the time-complexity of the CSP. Scenario 2 considers the rotation of some π-containers, denoted with a prime symbol ($'$) in Fig. 18.5. They have been rotated by 90° or 180° on the horizontal (or vertical) plane and their FLB corner coordinates are different of the sensor nodes. Table 18.1 gives for each scenario the FLB corner and sensor node coordinates.

For each scenario, the global neighbor graph (for a given transmission range R) can be processed by the wireless sensor nodes using the successive operations, described in Sect. 3.2. The objective of the CSP problem is to find all solutions that satisfy the neighborhood and dimensions constraints. To evaluate the impact of the transmission range in terms of the number of solutions and computational time, we used different values of R, between the smallest and the largest normalized dimensions of the unitary π-containers.

18.5.2 Results and Discussions

The CSP mathematical model was coded using MATLAB, running on a *quad-core Intel* 2.4 GHz processor with 8 GB of RAM. The numerical results are summarized in Table 18.2. It includes computational time (in seconds) and the number of possible solutions which satisfy all the constraints of the problem, for different values of R. It is observed that we only obtain one solution for scenarios 1 and 2 (Fig. 18.6a, b) when R is equal to 2.25 and 2.5, respectively. In both cases, these virtual 3D layouts reflect the real composition of the composite π-container defined in Fig. 18.5. A running time inferior to 5 s shows that our proposed method yields quick and satisfactory results.

Table 18.1 Coordinates of the FLB corners and sensor nodes for each scenario

	Container dimensions		FLB corner and sensor node coordinates			
	Scenarios 1 and 2	Scenario 3	Scenarios 1 and 3		Scenario 2	
πc_i	$(L_i \times W_i \times H_i)$		$(x_i, y_i, z_i) = (x_{ni}, y_{ni}, z_{ni})$		(x_i, y_i, z_i)	(x_{ni}, y_{ni}, z_{ni})
1	2 × 3 × 2	2 × 1 × 1	(2,0,0)	(0,0,0)	(2,0,0)	(4,0,0)
2	2 × 3 × 1	2 × 1 × 1	(0,0,0)	(2,0,0)	(0,0,0)	(0,0,0)
3	2 × 1 × 1	2 × 1 × 1	(0,0,1)	(0,1,0)	(0,0,1)	(0,0,1)
4	2 × 2 × 1	2 × 1 × 1	(0,1,1)	(2,1,0)	(0,1,1)	(2,1,1)
5	3 × 2 × 1	2 × 1 × 1	(0,0,2)	(0,2,0)	(0,0,2)	(3,0,3)
6	1 × 1 × 1	2 × 1 × 1	(0,2,2)	(2,2,0)	(0,2,2)	(0,2,2)
7	1 × 1 × 1	2 × 1 × 1	(1,2,2)	(0,0,1)	(1,2,2)	(2,3,3)
8	1 × 1 × 1	2 × 1 × 1	(2,2,2)	(2,0,1)	(2,2,2)	(2,2,2)
9	1 × 3 × 1	2 × 1 × 1	(3,0,2)	(0,1,1)	(3,0,2)	(4,3,2)
10		2 × 1 × 1		(2,1,1)		
11		2 × 1 × 1		(0,2,1)		
12		2 × 1 × 1		(2,2,1)		
13		2 × 1 × 1		(0,0,2)		
14		2 × 1 × 1		(2,0,2)		
15		2 × 1 × 1		(0,1,2)		
16		2 × 1 × 1		(2,1,2)		
17		2 × 1 × 1		(0,2,2)		
18		2 × 1 × 1		(2,2,2)		

Table 18.2 Computational results

	Number of solutions/time (s)		
R	Scenario 1	Scenario 2	Scenario 3
1	492/236	2776/575	720/5894
1.5	24/37	720/288	248/1985
2	2/0.8	12/8.6	**2/15.5**
2.25	**1/0.5**	4/7.1	2/2953
2.5	4/2.9	**1/4.9**	5688/18,232
3	180/75.8	322/187	–/–

For scenario 3, we obtained a minimum of two layouts (Fig. 18.6c, d) with $R = 2$ and a computational time inferior to 15 s. The time is satisfactory but the neighborhood information is not sufficient to determine the unique layout. Although Fig. 18.6c shows the real composition of the composite π-container, the same set of constraints is being respected with the obtained stacking Fig. 18.6d. This can be explained by the homogeneity set of π-containers and the orientation constraints which leads to a distribution of the sensor nodes as a regular grid. To overcome this problem, a

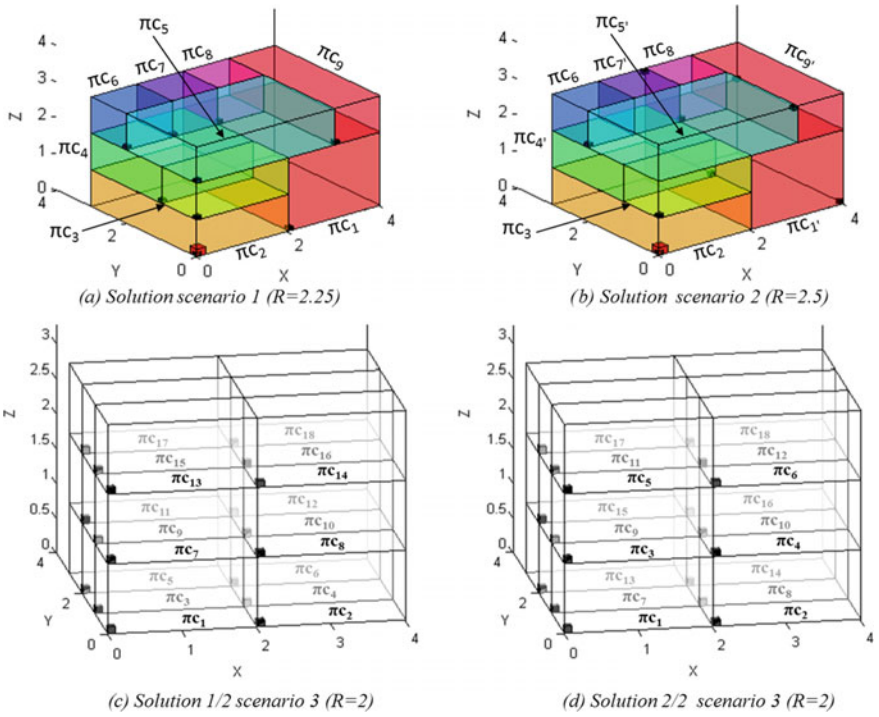


Fig. 18.6 Virtual 3D layout obtained for each scenario

domain filtering was developed to identify a unique layout. The successive operations (described in Sect. 3.2) are executed twice to obtain a new neighborhood graph. This can be achieved from the gateway node with a different transmission range, or by using a second gateway fixed to the composite π-container. The information is used to filter the space of possible solutions where each solution must fit with the two neighborhood graphs. For instance, a second gateway at the front-left-top corner of the composite π-container (scenario 3) can be used to generate new neighbor constraints between the gateway and the π-container 13. Hence the solution 2 (Fig. 18.6d) would be eliminated and the unique layout will be obtained without increasing the computational time

From Table 18.2, the computational time and the number of solutions increase with R close to 1 or 3. This can be explained when we observe the neighbor graphs resulting from the transmission range, as depicted Fig. 18.7. In the first case ($R=1$), the neighbor graphs are disconnected. A lot of vertices are isolated (with no edges connected) in scenario 1 and 2. Scenario 3 shows a disconnected graph with 2 connected components. Conversely, all vertices are strongly connected with a value of $R=3$. For these values of R, the neighborhood graphs are not relevant to reduce the research space, thereby increasing the number of solutions and the running time.

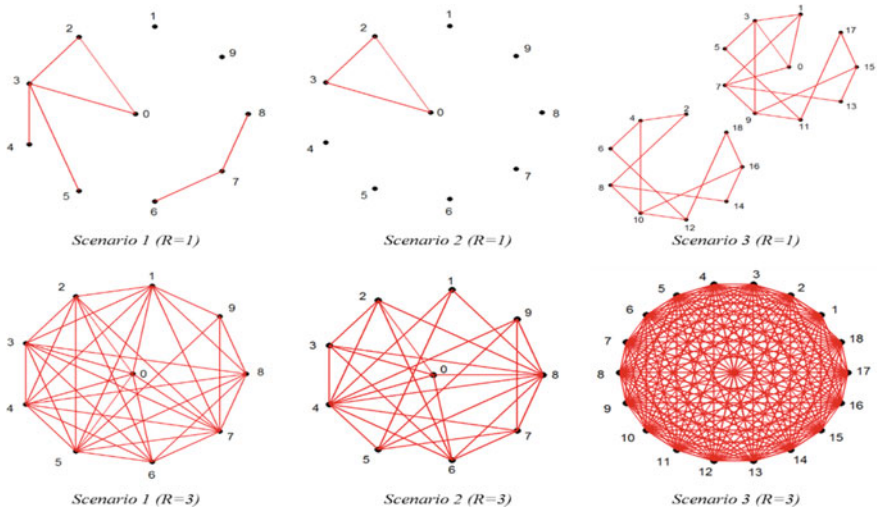


Fig. 18.7 Neighbor graphs obtained from $R=1$ and $R=3$

Therefore, the connectivity of the neighbor graph influences the solving of the CSP problem. Figure 18.8 illustrates the scaled connectivity and the connectivity properties (i.e., connected or disconnected) of the neighborhood graphs for a different transmission range. We can observe that the best results for each scenario are obtained with a minimum connected graph. This interesting result could be exploited in our approach to create pertinent and sufficient neighborhood constraints, and deduce the virtual 3D layout of the composite π-container. Before executing the CSP, the gateway could validate the relevance of the global neighbor graph and decide to replay the successive operations with a new transmission values (R) if needed.

18.6 Conclusion and Future Works

The Physical Internet is a highly dynamic transport and logistic system, where composite π-containers can be quickly set up and/or modify in order to exploit new opportunities leading to an increase in the sustainability of the global logistics chain. However, such a system requires a perfect synchronization between the physical and informational flows. The identification and localization of stacked π-containers in a composite π-container can help to maintain the traceability information, and thus contribute to the control of the Physical Internet

In this chapter, a π-container localization system was introduced to dynamically generate and maintain a virtual 3D layout of the composite π-container. The system is based on the use of wireless sensor network and the neighborhood relationships between nodes. It allows us to obtain a relative position between the π-containers.

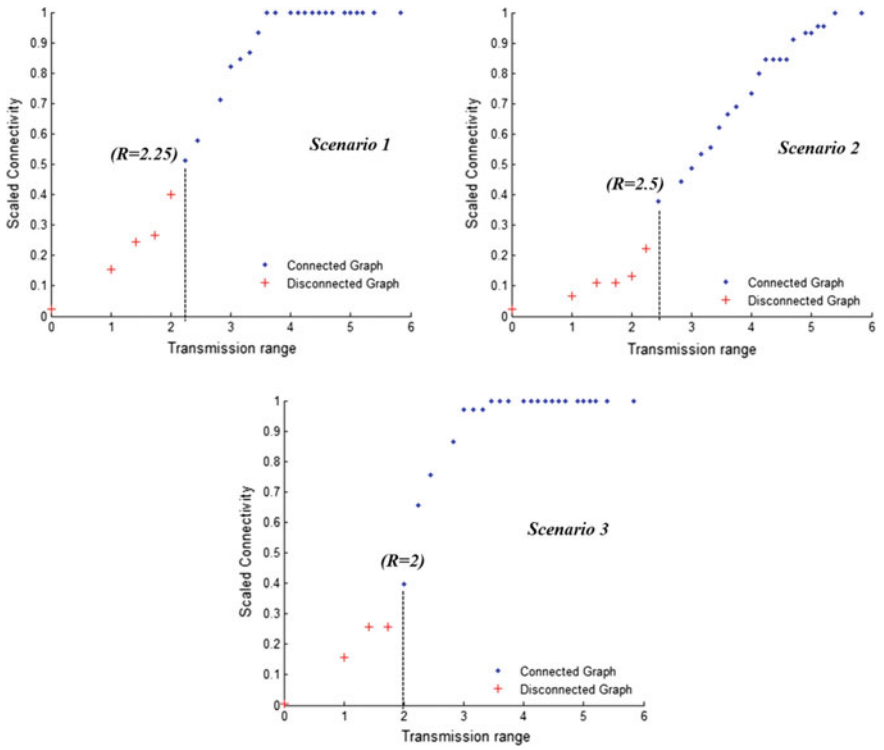


Fig. 18.8 Graph connectivity and transmission range

This information is used as a new type of constraints in a CSP problem where each solution represents a possible arrangement of the π-containers. As a proof-of-concept, the proposed approach and model have been validated through computational experiments. The results show that the virtual 3D layout reflecting the spatial distribution can be obtained within a reasonable amount of time when the neighborhood information is relevant and using domain filtering. This approach can be applied to provide up-to-date information (permanent inventory), as well as detect any error during the encapsulation process, which might have a negative impact on the overall effectiveness and efficiency of the Physical Internet. Furthermore, new value-added services such as guidance information for loading/unloading systems can be derived

Future works will focus on the graph theory to formally validate the relevance of the global neighbor graph. Although energy-efficient protocols for the discovery and forwarding mechanisms can be used, a study of the energy consumption of the sensor nodes will be also conducted. Finally, the opportunities to develop dynamic and decentralized CSPs will be also studied.

References

1. Abbate S, Avvenuti M, Corsini P, Vecchio A (2009) Localization of shipping containers in ports and terminals using wireless sensor networks. In: International conference on computational science and engineering, pp 587–592
2. Ballot E, Montreuil B, Meller R (2014) The physical internet: the network of the logistics networks. La Documentation Française, Paris
3. Chen C, Lee S, Shen Q (1995) An analytical model for the container loading problem. Eur J Oper Res 80(1):68–76
4. Hashimoto N, Isshiki N, Iguchi M, Morisaki M, Ishii K (2006) Assets location management solution based on the combination of SmartLocator and RFID. NEC Tech J 1:92–96
5. Li H, Zhou Y, Tian L, Wan C (2010) Design of a hybrid RFID/GPS-based terminal system in vehicular communications. In: 6th international conference on wireless communications networking and Mobile Computing (WiCOM), pp 1–4
6. Lin Y, Meller R, Ellis K, Thomas L, Lombardi B (2014) A decomposition-based approach for the selection of standardized modular containers. Int J Prod Res 52(15):4660–4672
7. McFarlane D, Sheffi Y (2003) The impact of automatic identification on supply chain operations. Int J Logist Manag 14(1):1–17
8. McKinnon A, Browne M, Whiteing A, Piecyk M (eds) (2015) Green logistics: improving the environmental sustainability of logistics. Kogan Page Publishers, UK
9. Meller R, Lin Y, Ellis K (2012) The impact of standardized metric physical internet containers on the shipping volume of manufacturers. In: Proceedings of 14th IFAC symposium on information control problems in manufacturing, Bucharest, Romania, pp 364–371
10. Montreuil B (2011) Toward a physical internet: meeting the global logistics sustainability grand challenge. Logist Res 3(2–3):71–87
11. Montreuil B, Ballot E, Tremblay W (2015) Modular design of physical internet transport, handling and packaging containers. In: Smith et al (ed) Progress in material handling research vol 13, MHI, Charlote, NC, USA, to appear
12. Montreuil B, Meller R, Ballot E (2010) Towards a physical internet: the impact on logistics facilities and material handling systems design and innovation. In: Gue K et al (eds) Progress in material handling research. Material Handling Industry of America, USA
13. Nekoogar F, Farid D (2012) Characteristics and limitations of conventional RFIDs. In: Ultra-wideband radio frequency identification systems. Springer, Boston, MA, pp 25–49
14. Röhrig C, Spieker S (2008) Tracking of transport vehicles for warehouse management using a wireless sensor network. In: International conference on intelligent robots and systems, pp 3260–3265
15. Sallez Y, Montreuil B, Ballot E (2015) On the activeness of physical internet containers. In: Borangiu T, Thomas A, Trentesaux D (eds) Service orientation in holonic and multi-agent manufacturing, vol 594. Springer, New York, pp 259–269
16. Sarraj R, Ballot E, Pan S, Montreuil B (2014) Analogies between internet network and logistics service networks: challenges involved in the interconnection. J Intell Manuf 25(6):1207–1219
17. Shavazi A, Abzari M, Mohammadzadeh A (2009) A research in relationship between ICT and SCM. Eng Technol 50:92–101
18. Spieker S, Röhrig C (2008) Localization of pallets in warehouses using wireless sensor networks. In: 16th mediterranean conference on control and automation, pp 1833–1838
19. Thiesse F, Dierkes M, Fleisch E (2006) LotTrack: RFID-based process control in the semiconductor industry. IEEE Pervasive Comput 5:47–53
20. Zhou J, Zhang H, Zhou H (2015) Localization of pallets in warehouses using passive RFID system. J Central South Univ 22(8):3017–3025

Chapter 19
An Information Framework of Internet of Things Services for Physical Internet

19.1 Introduction

The current logistics systems is still recognized to be unsustainable in terms of economy, environment, and society despite several innovative logistics paradigms accompanying projects have been proposed and undertaken to reserve the situation. The intelligent logistics exploiting the intelligence concepts thanks to the integration of advanced information and communication technology (ICR) to improve the overall efficiency significantly in terms of security, physical flow management, automation of processes. One of key factors of such logistics vision is intelligent freight-transportation system that uses the latest technologies, infrastructure, and services as well as the operation, planning, and control methods to efficiently transporting freight [1]. Such system is an effective solution to the significant transportation-related issues such as congestion, energy consumption, and environment. Simple Links[1] is an alternative framework of intelligent logistics. The project conducted by the consumer good forum aims to create, manage, and analyze digital links between items and their containers as a dynamic hierarchy, then to infer item-level tracking and monitoring information for any item based on searching the hierarchy for the best available track, monitor and prediction data at a given time. The objective of project is enabled by an underlying principle that make items and their containers communicate together to share information all along the logistics links by using low cost identification technology (i.e., QR code). The primary demonstration results of the project imply the positive impact and economical value end-to-end. To reduce the environment damage caused by inefficient ways of logistics activities, the Intelligent Cargo project (iCargo) project[2] funded by European Union is expected to provide a potential

[1]Simple Links project introduction, https://www.theconsumergoodsforum.com/wp-content/uploads/2018/02/CGF-Simple_Links_White_Paper-1.pdf.

[2]iCargo project, https://www.ec.europa.eu/digital-single-market/en/news/intelligent-cargo-more-efficient-greener-logistics.

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0_19

solution. Relying heavily on ICT, the iCargo is capable of self-identification, context detection, service access, status monitoring and registering, independent behavior and autonomous decision-making. Thus, the end-to-end visibility of shipment flow is improved by a cargo tracking system [2–4] capturing the real-time information relating to the location, time, and status of the vehicles and carried iCargo. In this way, the logistics processes such as asset management, routing path of freight can be optimized to avoid empty trucks or congestion. Another logistics visions termed as green logistics indicates methods that employ advanced technologies and innovated equipment to tackle the environment issues such as greenhouse gas (GHG) emissions, noise, and accidents mainly caused by inefficient logistics operations [5]. To make the global logistics green, the environment aspect is concerned as the top priority in all logistics operations in particular, transportation, ware-housing, and inventory operations [6]. These three activities are conducted in a series of researches and investigated in Operation Research (OR) model. For example, to minimize the GHG emissions by the transportation, the operation is performed by optimizing four significant choices, namely, mode choice (i.e., plane, ship, truck, rail, barge or pipelines), usage of inter-modal transport (i.e., types of containers), equipment choice (i.e., type and size of transportation unit) and fuel choice (e.g., gasoline, bio-fuels, electric, etc.). As the optimal selection of these handling equipment is set off, the minimum emission is obtained, thus the environment sustainability is resulted in.

Recently, the Physical Internet (PI, or π) is an emerging paradigm that is strongly expected to achieve simultaneously the three dimensions of sustainability: economy, environment, and society at global scale [7]. Conceptually, by taking the Digital Internet as a metaphor the PI is built towards an open global logistics system founded mainly on physical, digital and operation interconnectivity through a standard set of modular containers (π-containers), collaborative protocols (π-protocols), and smart interfaces (π-nodes, termed for container distribution centers) for increased efficiency and sustainability [8, 9]. Accordingly, the PI does not manipulate the physical goods directly but such π-containers that encapsulate the physical merchandise within them. These π-containers are world-standard, smart, and green and modular contains. Particularly, they are characterized by modularity and standardization worldwide in terms of dimensions, functions and fixtures. In the existing logistics and supply chain the diversity of brands and types of products with various sizes and weights leads to a nearly infinite range of different sizes of carton boxes. Therefore, generating efficient unit loads from such a high variance of cases is complicated and leads to inefficient space utilization at the pallet level and as a consequence also on a truck level. To release such issues, the container standardization concept of the PI aims to determine a limited set of modular container dimensions subject to all specific requirements [10]. Thus, having a small number of such containers would make it much easier for goods to be rolled out of π-nodes and onto π-trucks and π-trains. In addition, as the open global logistics network, the PI allows all stakeholders share the infrastructure including the π-containers, π-nodes, and π-movers (denoted

for material handling equipment) to maximize its operation efficiency. In addition, the π-nodes and π-movers are designed and innovated to exploit as best as possible the standard and modular encapsulation. For example, the physical dimensions of π-movers such as π-pallets or π-trucks should be modular and can be adapted to fit any size of composite π-containers [11]. In this way, facilitation of logistics processes such as moving, storing, handling, transporting the unit loads enables the system to gain huge efficiency and sustainability. The work in [12] proved that container standardization increases the space utilization of trucks and material handling equipment. In the similar research, an evaluation of PI performance introduced in [13] indicated that transporting shipments in the PI network contributes to reduce the inventory cost and the total logistics system thanks to the consolidation facilitation of such standardized containers. Furthermore, such advantage characteristic of the PI would enable the PI network to achieve more sustainable environment since the shipment consolidation approach was proved to mitigate the carbon and energy waste effectively in the traditional logistics system [14, 15].

A project has been conducted by CELDi (Center for Excellence in Logistics and Distribution) to examine more comprehensive impact of the PI on the performance of existing logistics system in the U.S.A. The primary demonstration results declared in [9] imply various positive impacts of the PI on all three sustainable aspects. Concretely, if the PI was rolled out on 25% of flows in the U.S.A, its impact would represent a saving of 100 billion USD, a reduction of 200 million tons of CO_2 emissions, and a decrease of 75% in the turnover of long distance for vehicle drivers driving heavy goods. Another project led within CIRRELT (Inter-university Research Centre on Enterprise Network, Logistics and Transportation) is to estimate the potential energy, environmental and financial gains for one industrial producer of manufactured goods exploiting an open distribution network. The simulation results from several scenarios introduced in [9] show the impact of PI with manifold gains including an increase in the fill rate of vehicles (e.g., trucks, trains, etc.), energy savings (saving millions of liters of diesel per year), and a reduction of overall logistics cost.

The PI is a long-term vision for an end-to-end global logistics network, and several alliances like MHI[3] and ALICE[4] have already adopted and then promoted the PI conceptualizations and practices. They also have decided to declare the PI as the ultimate logistics goal and set up a comprehensive roadmap to realize the PI concept by 2050. To reach the full-fledged PI many factors need to be taken into account simultaneously, including physical objects such as π-containers, π-movers, and π-nodes as well as informal abstracts including π-protocols and the Physical Internet management systems (PIMS). The PI will thus lead to the development of an open logistics system for connecting the physical objects to the global Internet.

[3]MHI, the largest material handling logistics, and supply chain association in the U.S., has created a community of industry thought leaders called the U.S. Roadmap for Material Handling and Logistics. http://www.mhi.org/.

[4]ALICE (Alliance for Logistics Innovation through Collaboration in Europe), http://www.etp-logistics.eu/.

This concept suggests significant organizational evolutions, in which the logistics objects are expected to be intelligent and autonomous since they are a communication channel and a stock of information at the same time. In this context, the PI is a key player poised to benefit from the Internet of Things (IoT) [16] revolution since millions of π-containers are moved, tracked, and stored by a variety of the π-nodes and π-movers each day. By embedding intelligent capabilities such as sensing, communication, and data processing into the PI components, IoT enables seamless interconnection of the heterogeneous devices and to get complete operational visibility and allow for the best real-time decisions in the logistics processes. For example, IoT enables managers to monitor the performance of machines, ambient conditions, energy consumption, status of inventory, or the flow of materials. Thus the benefits are contingent upon the design of IoT architecture for the PI and particularly the IoT services ubiquitously to manage and control efficiently the industrial automation processes in the logistics domain. However, much of the PI and IoT literature, to date, has been largely disjointed without much emphasis on theory or practical applicability [17]. This chapter examines the trends of IoT applications in the PI and uncover various issues that must be addressed to transform logistics technologies through the IoT innovation. In this regard, the main contributions of chapter are summarized as follows:

- The PI paradigm accompanying the state-of-the-art developments of related projects are highlighted. In addition, the proposition design for the components of PI is introduced toward IoT application in the PI. Particularly, an active distributed system is proposed for the PIMS to enable smooth flows of the π-containers through the π-nodes.
- An information framework is proposed based on exploiting the IoT embedded in the physical devices of the PI and their active interaction.
- A service-oriented architecture (SOA) is proposed and described for IoT applied for the PI. Such SOA then exploits the information framework to create and deliver IoT services for the PIMS.
- By adapting the proposed framework, a case study aiming at developing an IoT service is presented to illustrate an efficient management service of logistics operations in the PI.

The rest of the chapter is organized as follows: Sect. 19.2 describes the key infrastructure enabling the PI and our proposition design to create an IoT in the PI network. In Sect. 19.3, we propose an SOA, which is designed to adapt to the logistics environment to provide the IoT logistic services. After that, Sect. 19.4 introduces a case study used the proposed IoT architecture as well as the proposed SOA architecture to create and deliver a value-added services for the PI management. Finally, Sect. 19.5 concludes the chapter and proposes further developments.

19.2 IOT Infrastructure for Physical Internet

In this section, we will present the key elements including π-containers, π-nodes, and π-movers as well as the state-of-the-art proposition design that enables the IoT for the Physical Internet.

19.2.1 π-Containers

The π-containers are a key element enabling the success of the PI operation. The π-containers are classified into three functional categories: transport, handling, and packaging containers (termed as T/H/P-containers respectively) with their corresponding modular dimensions [18]. The propositions are introduced in [19] to design the modular sizes of these π-containers. Although the set of modular dimensions must be subject to an international standard committee, the partners of the CELDi PI project [12, 20] developed a mathematical model to determine the best container size to maximize space utilization. In parallel, the project MODULUSHCA[5] funded by the 7th framework Program of the European commission is the first project to design and develop π-containers (termed as M-boxes in this project) dedicated for fast-moving consumer goods (FMCG). Generally, the FMCG includes daily used goods of consumers with wide range of small/medium sizes such as pharmaceutical, consumer electronic, personal care, household care, branded and packaged foods, spirits, and tobacco. The M-boxes are sized and designed by the methodological engineering process introduced in the research [10]. They demonstrated that a set of external dimensions of π-containers including the following values in meters {0.12, 0.24, 0.36, 0.48, 0.6, 1.2, 2.4, 3.6, 4.8, 6, 12} increases the space utilization at the different unit load levels [11]. With the modularity and interlocking structure, the encapsulation concept [8] is applied to create efficient unit loads that facilitate material handling processes such as moving, loading, or storage. Figure 19.1 illustrates such concept realized by composing nine smaller π-containers to create a composite π-container as an efficient unit load.

Note that, the physical encapsulation is applied in the three types of π-containers. Accordingly, a P-container containing directly several passive goods is places in a H-container, itself contained in a T-container. Indeed, a number of smaller π-containers are optimized such as their composition block (composite π-container) fits perfectly with the internal space of a large π-container.

Additionally, the PI emphasizes the importance of informational and communicational encapsulation. This is achieved by applying IoT and embedding the accompanying technologies such as RFID, wireless sensor networks (WSNs). With such ICT integration, the π-containers become smart IoT objects [21] having basic capabilities such as identification, ambient sensing, computation, and communication. Since the π-containers are manipulated worldwide by all stakeholders, the information relating

[5]MODULUSHCA project, www.MODULUSHCA.com.

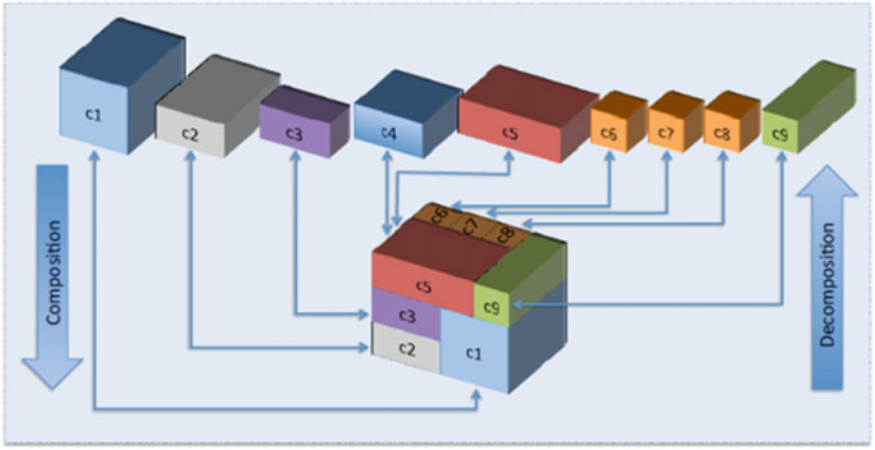


Fig. 19.1 An example of encapsulation concept that composes nine π-containers to create an efficient unit load [23]

to their physical status and context must be captured, coded, protected, and transferred accurately. The EPC global standard provides a solution that allows identifying π-containers uniquely. Concretely, the container code is created in line of RFID tag with the EPC standard. Thus, when the code is put into service, the information fed to an EPC Information Service (EPCIS)[6] is shared among stakeholders. Meanwhile, the sensing capability of the sensor node allows the ambient environment condition of the container to be monitored periodically. With such ability, the π-container is able to identify its state and report it, compare its state with the desired one, and send information (e.g., warning) when certain conditions are met. Furthermore, integrated with a memory, the sensor node can store and maintain relevant data. Applying for the PI network, since π-containers are equivalent to the data packets flowed in the Digital Internet networks, information relating to fundamental specification of π-container should be stored in the memory such as container identifier, container dimension, container category. In addition, to support routing protocols effectively, routing information (i.e. previous/next/final destination address) of the π-container provided by the PIMS is also added to the memory of the sensor. More important, any problem in the logistics process along the supply chain will be recorded in the wireless sensor node memory, so such information can be checked using any device such as a computer, tablet or even a mobile phone. Processing capability enables the nodes to perform specific tasks and provide corresponding additional functionalities for the containers. Meanwhile, enabled by integrated wireless transceivers, the π-container can communicate with the different support systems (e.g., manufacturing systems, supply chains, maintenance systems, PIMS) and other active π-containers to enable IoT as well as IOT services [22]. Practically, an example of integration of

[6]EPCIS, GS1 standard, https://www.gs1.org/epcis/epcis/1-1.

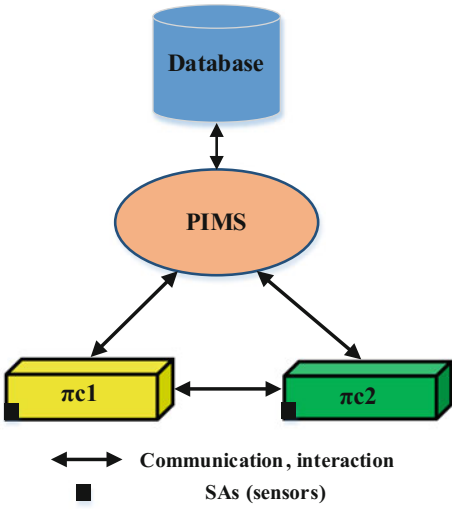


Fig. 19.2 The wireless sensors or smart tags embedded in the π-containers as their agents enable interaction between them and with a PIMS actively

all these capabilities in the Intelligent Container called InBin[7] recently developed by the Fraunhofer Institute for Material Flow and Logistics to support transporting the perishable food products efficiently. Another project named TRAXENS[8] has been developed smart multi-modal containers (i.e., equivalent T-containers in PI) that can achieve a huge gain in efficiency, service, and protection of the planet since the visibility of the cargo containers are obtained in real time.

With such ICT integration, each smart π-container is represented by a corresponding intelligent agent (e.g., a smart tag, wireless sensor nod, etc.) as illustrated in Fig. 19.2. Each agent is as a communication channel enabled by wireless communication technologies and holds a stock of significant information relating to the products and their status. In addition, it helps ensure the identification, integrity, routing, conditioning, monitoring, traceability, and security of each π-container. It also enables distributed handling, storage, and routing automation [23]. Throughout the PI, stakeholders can access necessary data by interacting with the agents. However, based on the roles of the requesters more restricted data maybe required and accessed to ensure the security and privacy of data. A Modulusca common data model proposed in [24] is composed of four data types: business data, shipment data, network data, and public data, which can be accessed by corresponding actors with the granted right to support exchanging information among the partners of the PI.

[7] InBin project, http://www.industrie40.iml.fraunhofer.de/en/ergebnisse/inbin.htm.
[8] TRAXENS project, http://www.traxens.com/en/.

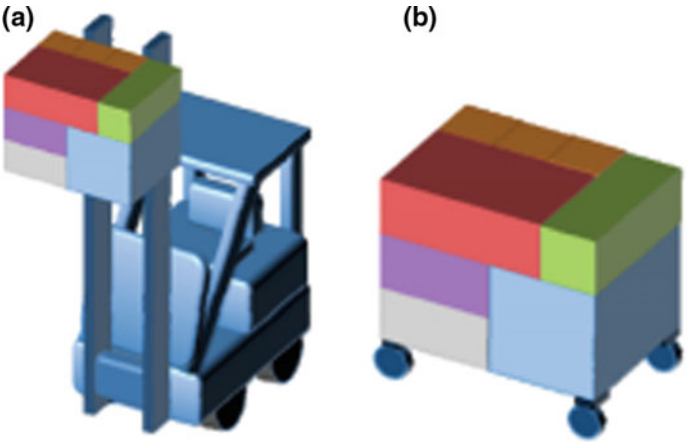


Fig. 19.3 Conceptual illustration of π-movers designed to exploit the modularity of π-containers. **a** π-truck-lift. **b** π-mover

19.2.2 π-Movers

In the Physical Internet, π-containers are generically moved around by π-movers. Moving is used here as a generic equivalent to verb such as transporting, conveying, handling, lifting, and manipulating. The main types of π-movers include π-transporters, π-conveyors, and π-handlers. The latter are humans that are qualified for moving π-containers. All π-movers may temporarily store π-containers even though this is not their primary mission.

Since the π-movers manipulates directly with the π-containers, they are designed to exploit as best as possible the characteristics of π-containers. From physical perspective, dimensions of π-movers are innovated to subject to the modularity standard of π-container so as moving such containers or composite π-containers with different sizes are facilitated. Figure 19.3 illustrates the conceptual model of π-movers designed to exploit the modularity of π-containers.

Regarding to information perspective, the widespread adoption of IoT technologies enables smart inventory and asset management. In particular, π-movers can act as active agents, which can interact with π-containers and the PIMS for information sharing and manage them temporarily in the distributed manner. Thus, through the interaction, the basic information includes the specification of π-containers such as ID, dimensions, final destination. In addition, the status of π-container and also π-movers are monitored in real time. For example, the PIMS can be alerted when a π-truck is being over-utilized or when an idle π-pallet should be assigned to do other task. Figure 19.4 illustrates an example showing the activeness of π-movers (π-pallet (πP) and π-truck (πT)) enabled by equipped wireless sensors or gateway.

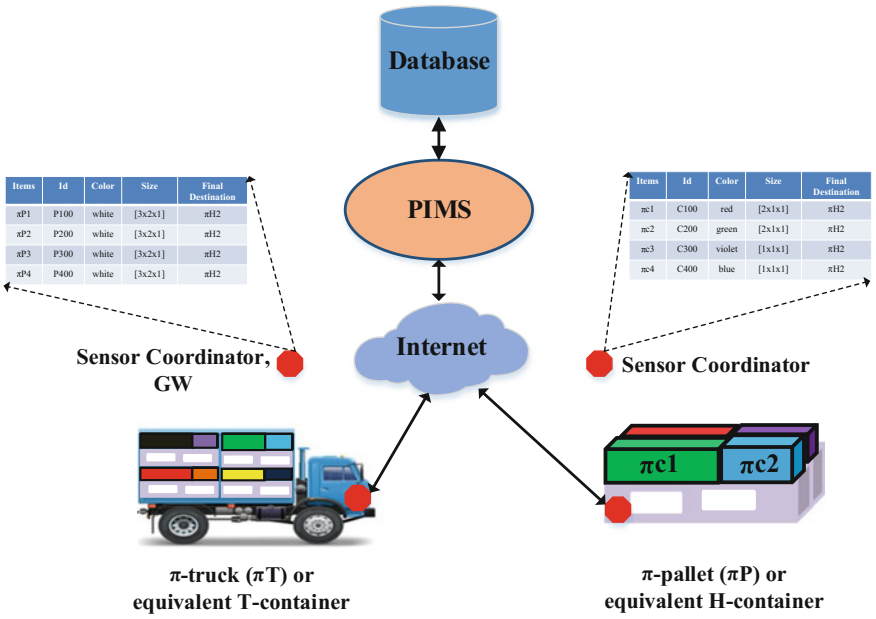


Fig. 19.4 IoT logistics services provided by IoT infrastructure of physical internet like active πT and πP enabled by equipped wireless sensors or gateway

19.2.3 π-Nodes

In the PI vision, the π-nodes are locations playing a role as smart interfaces to enable realizing universal interconnectivity at the operational level. For example, π-gateways enable efficient and controlled entry of π-containers into the PI as well as their exit from the PI. Generally, a π-node includes π-movers and/or other embedded π-nodes permanently or temporarily, which are collaborated for joint purposes of material handling (e.g., composing π-containers, moving composite π-containers, storing composite π-containers, etc.). Table 19.1 summarizes the key π-nodes designed for the PI network [23]. As container distribution centers almost logistics activities are taken place and transformed here dynamically. Thus the π-nodes must be designed so that they exploit as best as possible the characteristics of π-containers to support smooth movement of the containers. In the next section, an active distributed PIMS for these smart interfaces are proposed to enable the smooth physical flows of smart π-containers.

Table 19.1 Key π-nodes with their specific functionality in the PI

π-node	Functionality
π-transits	Transferring π-carriers from inbound π-vehicles to outbound π-vehicles
π-switch	Unimodal transfer of π-containers from an incoming π-mover to a departing π-mover
π-bridge	One-to-one multimodal transfer of π-containers from an incoming π-mover to a departing π-mover
π-sorter	Receiving π-containers from one or multiple entry points and having to sort them so as to ship each of them from a specified exit point, potentially in a specified order
π-composer	Constructing composite π-containers from specified sets of π-containers
π-store	Storing π-containers within a specific time duration
π-gateway	Receive π-containers and release them so they and their content can be accessed in a private network
π-hub	Transfer of π-containers from incoming π-movers to outgoing π-containers

19.2.4 Active Distributed PIMS for π-Nodes

Due to the high dynamic and structural complexity of the PI networks, central planning and control of logistics processes become increasingly difficult. The difficulty is amplified by the need of efficient management of a huge number of π-containers and logistics assets in each π-node. Therefore, distributed and autonomous control and management of logistics processes are required in the context of the PI. In other words, the PIMS should be designed in distributed manner enabling management of the facilities and logistics assets efficiently. In addition, the PIMS should be active to exploit as best as possible the capabilities of IoT π-facilities. The term "activeness" refers to the ability of PIMS in monitoring status of the logistics assets in real time and based on this information scheduling, processes are planned flexibly and effectively. In this way, the utilization of assets can be optimized. This section describes the proposed active distributed system designed to enable IoT for PI in both information and physical flow perspectives.

With the active distributed system, the logistics processes are monitored and controlled effectively. Thus, intention mistakes or errors can be traced and the exact processes causing such issues can be found thank to the activeness of PIMS to correct them. Generally, the main mission of π-nodes is to ensure the π-container transferring efficiently and sustainability from their inbound π-movers carrying π-containers to outbound π-mover. In addition, enabled by IoT, the logistics assets

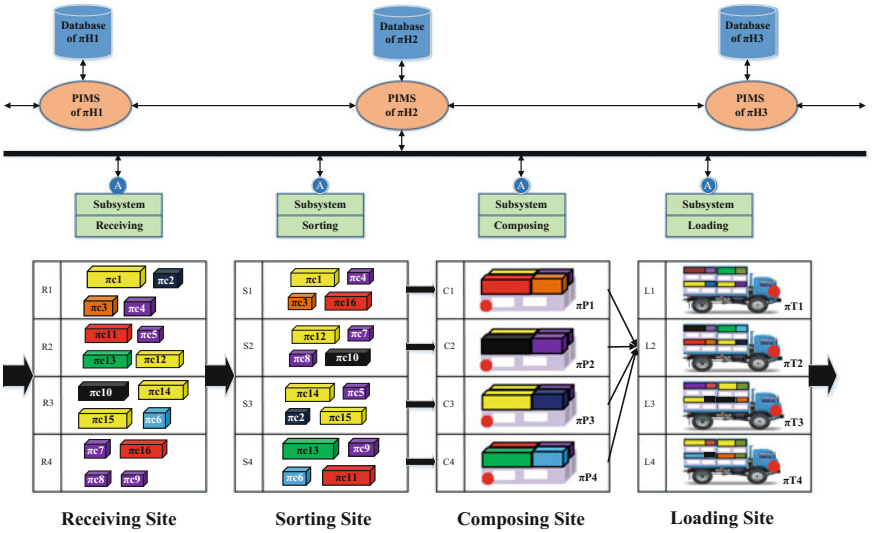


Fig. 19.5 A typical distributed PIMS to manage incoming and outgoing π-containers and the logistics assets in a πH2

can connect to the PIMS. Therefore, all the π-nodes can manage not only their incoming and outgoing π-containers participating in the logistics processes but also their logistics assets in real time.

Figure 19.5 illustrates such a PIMS at a π-hub distributed into active subsystems, which are responsible for single functions or corresponding services.

In the same way, a subsystem evolving specific π-containers and facilities assigned by the PIMS manages and takes its function. An active subsystem is responsible for managing a set of PI assets, inventories (i.e., π-containers, π-movers) within a scheduled period to complete its specific task.

As defined in Table 19.1 the exchange of π-containers from carriers to another is the core activity of the π-hub (πH). The following demonstrates such exchange process in both physical and information flow with IoT integration. As shown in Fig. 19.5, the logistics process in the πH2 is divided in four sub-processes managed by four corresponding active subsystems.

19.2.4.1 Receiving

At the receiving site, all inbound π-movers, and π-containers are registered, verified, and scanned for validation and security purposes. At this stage, the activeness of π-facilities is exploited to cross-check the current incoming inputs against the information transmitted from πH1. After passing such initial processes, π-containers are directed to appropriate locations for their next involved processes. Usually, mov-

Table 19.2 Time and space schedule for π-containers at receiving site

Status of π-containers at receiving site					
πc	Moved by (π-mover)	Location (lane)	Outgoing time(min, max)	Next process	Next Hub
πc1	π-conveyor	R1	(07:00, 07:02)	Sorting	πH3
πc2	π-conveyor	R1	(07:00, 07:02)	Sorting	πH4
πc3	π-conveyor	R1	(07:00, 07:02)	Sorting	πH3
πc4	π-conveyor	R1	(07:00, 07:02)	Sorting	πH3
πc5	π-conveyor	R2	(07:00, 07:02)	Sorting	πH4

ing the π-containers is supported by π-conveyors or π-carriers depending on their next destination and next processes. To avoid collisions and balance the loads, the π-containers are allocated to go in different input lanes. For example, as illustrated in Fig. 19.5, every four π-containers are scheduled and moved in one lane (i.e., R1, R2, R3, R4) to ensure the smooth flow time window and space window. Table 19.2 illustrates a detail status of π-containers scheduled by the receiving subsystem at the receiving site. The schedule is sent to the sorting subsystem that is responsible for sorting the listed π-containers.

19.2.4.2 Sorting

Going out from the receiving site, the received and selected π-containers go to the sorting site which aims to sort such π-containers according to some rules as followings

- π-containers have the same final destination,
- or they may have different final destination but are scheduled to transit in the same next π-node.

Table 19.3 illustrates an example of status and schedule of π-containers at the sorting site.

In this example, π-containers, πc1 πc3, and πc4 having the final destination (πH3) are sorted and moved in the lane S1. Meanwhile, πc2 and πc5 are moved in the lane S3 since their final destination is πH4. As mentioned before, sorting mission is achieved by a π-sorter that incorporates a network of π-conveyors and/or other embedded π-sorters.

19.2.4.3 Composing

At the composing site, the sorted π-containers are composed to be composite π-containers, which are efficient loads with different sizes fitting the sizes of

Table 19.3 Time and space schedule for π-containers at sorting site

Status of π-containers at sorting site					
πc	Moved by (π-mover)	Location (lane)	Outgoing time(min, max)	Next process	Next Hub
πc1	π-conveyor	S1	(07:03, 07:04)	Composing	πH3
πc2	π-conveyor	S3	(07:03, 07:04)	Composing	πH4
πc3	π-conveyor	S1	(07:03, 07:04)	Composing	πH3
πc4	π-conveyor	S1	(07:03, 07:04)	Composing	πH3
πc5	π-conveyor	S3	(07:03, 07:04)	Composing	πH4

Table 19.4 Time and space schedule for π-containers at composing site

Status of π-containers at composing site					
πc	Moved by (π-mover)	Location (lane)	Outgoing time(min, max)	Next process	Next Hub
πc1	πP1	C1	(07:04, 07:05)	Loading	πH3
πc2	πP3	C3	(07:04, 07:05)	Loading	πH4
πc3	πP1	C1	(07:04, 07:05)	Loading	πH3
πc4	πP1	C1	(07:04, 07:05)	Loading	πH3
πc5	πP3	C3	(07:04, 07:05)	Loading	πH4

H/T-containers. In addition, their composing orders must be taken into account since some π-containers can be decomposed in the next π-nodes, thus they should be placed at the outermost locations of the composite π-container. In other words, any mistake in allocating π-containers can lead to the inefficiency of following logistics process. Generally, the composing task is completed by π-composers in combination with the pre-assigned π-movers such as π-pallets or larger π-containers to carry.

Table 19.4 illustrates an example of status and schedule of π-containers at the composing site.

After π-containers are composed to be a composite π-container, the information encapsulation is applied. Accordingly, the coordinator or gateway embedding into the π-pallet is responsible for managing their hold π-containers directly instead of the composing subsystem. This distributed management allows the PIMS reduce the management and storage cost of the related data.

19.2.4.4 Loading

The loading process deals with composite π-container or H/T-container to load them into π-transporters and then to move them to next π-node. Similar to the composing process, the loading process must follow significant rule as following:

Table 19.5 Time and space schedule for π-containers at loading site

Status of π-containers at loading site					
πc	Moved by (π-mover)	Location (lane)	Outgoing time(min, max)	Next process	Next Hub
πP1	πT2	L2	(07:05, 07:06)	Transporting	πH3
πP2	πT2	L2	(07:05, 07:06)	Transporting	πH3
πP3	πT3	L2	(07:05, 07:06)	Transporting	πH3
πP4	πT4	L2	(07:05, 07:06)	Transporting	πH3

– Number of composite π-containers including H-containers are optimized so as their spatial arrangement in a π-transporter maximizes the space utilization,
– Since the composite π-containers may have different final destination but same next destination, their loading orders must be taken into account to facilitate the unloading process in the next destination. In this case, the last-in-first-out (LIFO) rule is applied.

Table 19.5 illustrates an example of status and schedule of π-containers at the composing site.

In the example, the four π-pallets are moved and converged at the lane L2 of the loading site, at which they are loaded into the πT2 pre-assigned. On the informational perspective, the coordinator or gateway or IoT device mounted in the πT2 temporarily manages their loaded π-containers until they are handled in the next process.

19.3 Service-Oriented Architecture for the IOT

Since no single consensus on architecture for IoT is agreed universally, different architectures have been proposed by researchers [25]. With a numerous number of things moved dynamically in the PI scenarios, an adaptive architecture is needed to help devices dynamically interact with other things in real-time manner. In addition, the decentralized and heterogeneous nature of IoT requires that the architecture provides IoT-efficient event-driven capability as well as on demand services. In addition, the PIMS is difficult to implement and maintain at global scale due to the lack of an efficient, reliable, standardized, and low-cost architecture. Furthermore, the ever-changing demands of businesses and the vastly different needs of different end users should be met by providing customization functionality to the end users and organizations under a flexible SOA. Thus an SOA is considered an efficient method achieve interoperability between heterogeneous devices in a multitude of way [16, 26, 27]. In addition, SOA is considered suitable for such demand-driven logistics chains. In particular, SOA can integrate logistics processes and information; and sharing such information can help create a better environment for real-time collab-

oration, real-time synchronization, and real-time visibility across the entire logistics chain.

This section describes the proposed SOA for creating the IoT logistics services in PI. The architecture comprises four layers: physical layer, network layer, service layer, and interface layer. In the following subsections, these layers are presented to adapt to the IoT of PI.

19.3.1 Physical Layer

The physical layer involves perceiving the physical characteristics of things or surrounding environment. This process is enabled by several identification and sensing technologies such as RFID, WSN [28–30]. For example, by embedding an intelligent sensor to a π-container, the environment condition (e.g., temperature, humidity, etc.) around it can be sensed and monitored in real time. In addition, since communication is enabled by these sensors, the π-containers can exchange information and identify each other.

In addition, this layer is in charge of converting the information to digital signals, which are more convenient for network transmission. However, some objects might not be perceived directly. Thus, microchips will be appended to these objects to enhance them with sensing and even processing capabilities. Indeed, nanotechnologies [31], communicating material [32, 33] and embedded intelligence will play a key role in the physical layer. The first one will make chips small enough to be implanted into the objects used in our everyday life. The second one will enhance them with processing capabilities that are required by any future application.

At the lowest layer of the architecture, the physical layer provides sets of information periodically or in passive mode. With the massive logistics activities, the information levels should be categorized and standardized. For example, the information used to realize the four classes of activeness of π-containers can be classified into four corresponding levels [22]:

- Passive information: it is collected from static or dynamic data stored in the RFID tag or sensors. Such information relates to π-container specification and location for providing tracking and traceability function.
- Triggering information: This information is perceived from sensing and detecting by adequate sensors. Therefore, detected problems are sent to the PIMS as alert message. Such information provides the monitoring function.
- Decisional process information: This information is obtained through the interaction and communication among proximity π-containers. The management of incompatibility between π-containers is an example of services served by such information.
- Self-organized information: The active π-containers are self-sufficient and able to provide services based on the information obtained from the π-infrastructure in the previous class.

Table 19.6 Design considerations for IIoT applications (adapted from [27])

Design goals	Description
Energy	How long can an IoT device operate with limited power supply?
Latency	How much time is need for message propagation and processing?
Throughput	What is the maximum amount of data that can be transported through the network?
Scalability	How many devices are supported?
Topology	Who must communicated with whom?
Security & safety	How secure and safe is the application?

19.3.2 Network Layer

The role of network layer is to connect all heterogeneous things together and allow them to share the information with other connected things [16]. In addition, the networking layer is capable of aggregating information from existing IT infrastructures supporting the logistics processes (i.e., π-movers, π-facilities). In SOA-IoT, services provided by IoT or a collective group of devices are typically deployed in a heterogeneous network and all related things are brought into the service Internet [16] for further accessing and sharing. Since, the networking layer mainly provides information collected from the physical layer to the service layer, QoS management, service discovery and retrieval, data and signal processing, security, and privacy according to the requirements of users/applications are some significant issues [34]. On the other hand, the dynamic changing of network topology due to leaving or joining of IoT devices may lead to non-robust of the network. Practically, since the network is the backbone to realize the IoT, designing it must consider the following significant challenges listed in Table 19.6.

19.3.3 Service Layer

This layer relies on middle-wares to integrate multiple source of heterogeneous information provided by heterogeneous IoT devices to create valuable services. These services in turn are exploited to support the logistics process of the active subsystem as well as end users to monitor or track their orders. This layer is responsible for identifying and realizing services that could utilize the IoT infrastructure to service operations in the logistics service. Basic traceability or tracking functions, monitoring, scheduling, routing are typical services in the logistics operations. In addition, all service-oriented issues including information exchange and storage, database management, search engines (database search, service search), and communication protocols among services are resolve by the layer [16, 26].

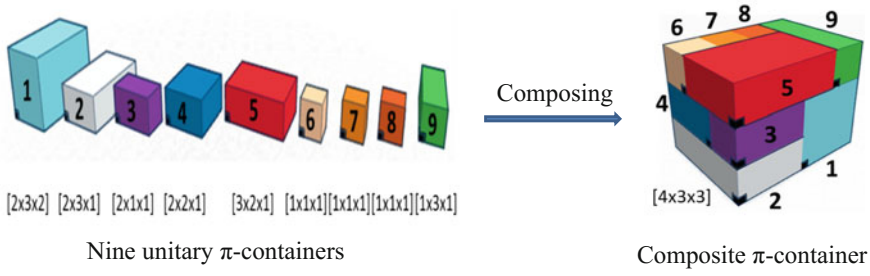
19.3.4 Interface Layer

In IoT, since a large number of IoT devices are made by different manufactures/vendors and they do not always follow the same standards/protocols, there are many interaction problems with information exchange, communication between things, and cooperative event processing among different things. Furthermore, the constant increase of IoT objects participating in an IoT makes it harder to dynamically connect, communicate, disconnect, and operate. Therefore, the mission of the interface layer is to enable the services to be accessed and used by managers or end users. Web Services [34] are technologies that integrate a set of standards and protocols to exchange data between applications developed in different programming languages and they can run on any platform. Therefore, the Web Services can be used to exchange data in both private IoT devices or private networks and the Internet. Interoperability is achieved by open standards proposed by organizations such as OASIS and W3C.

19.4 Management of Composite π-Containers: A Case Study

Composite π-containers composed of specified sets of smaller π-containers are efficient unit loads, which are absolutely key in improving transport, storage and handling efficiency across the PI network. By exploiting the modularity and interlocking structure of the π-containers, the unit loads have different sizes that are adapted to fit with the different sizes of material handling systems such as π-pallets (see Fig. 19.6). Next, such sets of π-pallets, which in turn, are loaded into a π-truck such that the utilized space of the truck is maximized prior to transport as illustrated in Fig. 19.5.

Because the exchange of π-containers is the core activities taken place continuously in the PI, a high frequency of transformation processes can introduce a desynchronization between the physical and information flows of the π-containers



Nine unitary π-containers

Composite π-container

Fig. 19.6 An efficient unit load (i.e. a composite π-container) is formed by composing nine unitary π-containers appropriately [37]

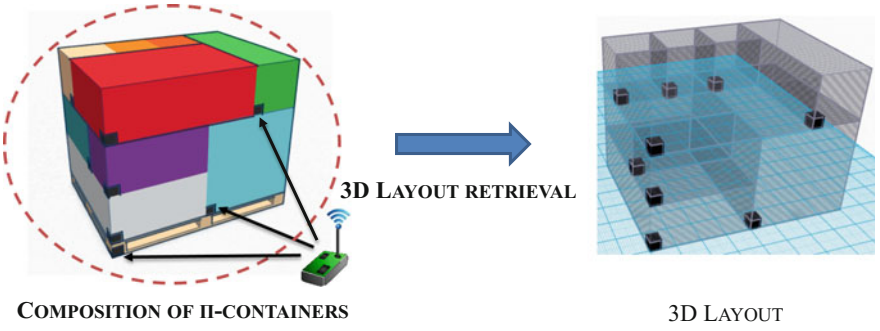


Fig. 19.7 3D layout retrieval from a composite π-container

maintained and previously stored in the PIMS. For example, an unexpected π-container can be placed on a π-pallet during the composing process or in a π-truck during the loading process. This type of issue leads to inefficiency in management of logistics processes because it requires additional costs and delay in recomposing or reloading activities. In addition, in reserve processes (i.e., decomposing or unloading) the part of unitary π-containers can be decomposed (i.e., de-palletized) automatically at the final destination or for order picking at the next destination. Such a process is facilitated of the position and orientation of the composed π-containers is available.

As a case study of utilizing the IoT technology, a methodology has been developed to obtain 3D layouts of composite containers, which is used for monitoring and validating such unit loads and addressing the above limitations [35]. This section presents the description of the approach as shown in Fig. 19.7, which is developed to provide these kinds of important information. Particularly, based on this layout, value-added services enabled by the IoT can be developed and used to enhance the efficiency of other logistics operations.

19.4.1 Architecture

The proposed network architecture of IoT shown in Fig. 19.8 contains both fixed and mobile infrastructure.

Applying the proposed architecture, the four-layer IoT architecture consists of the following elements. At the sensing layer, each sensor node equipped with each π-container contains information related to the specification of its container as well as the contained shipments. In addition, ambient conditions such as temperature, humidity are sensed and stored in the sensors. A gateway (GW) placed at a corner of π-pallet is responsible for coordinating the network formed by those sensors and managing the composed π-containers. At the networking layer, the sensors and the GW communicate on IEEE 802.15.4 links [36] to form an ad hoc network

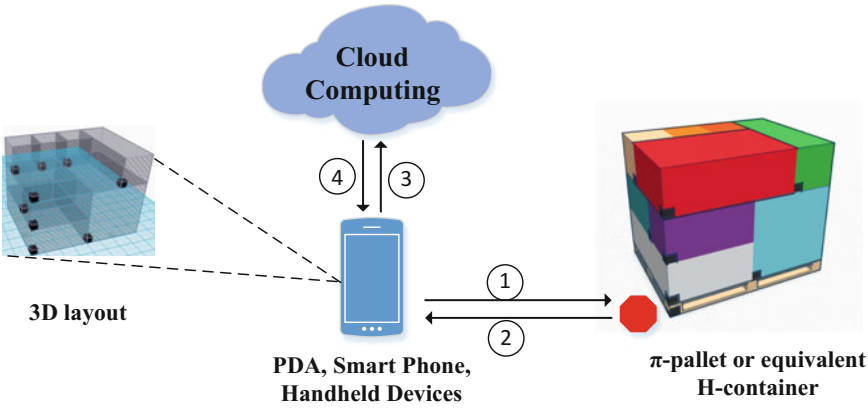


Fig. 19.8 IoT architecture for 3D layout retrieval at H-container level

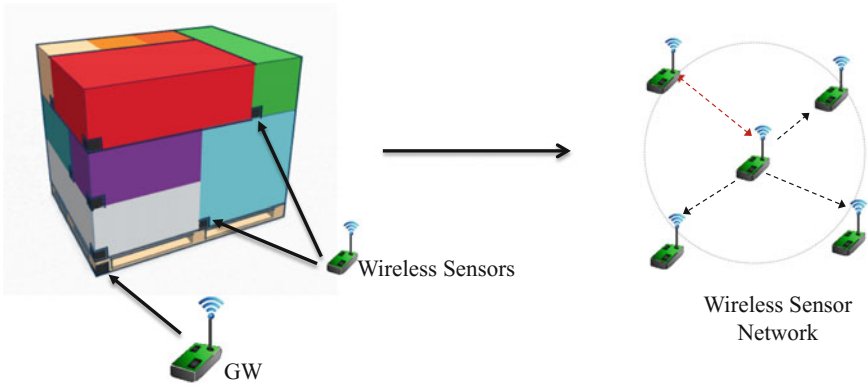


Fig. 19.9 At the networking layer, an ad hoc network is formed by nine wireless sensors and coordinated by the GW

coordinated by the GW (see Fig. 19.9). The mobile infrastructure usually includes PDA, smart phones or handled devices. These IoT devices request information from the GW/coordinator (flow 1 in Fig. 19.8). After receiving the demanded data (flow 2), it then sends it to the cloud for requesting the services (flow 3). Finally, the devices receive the services and display the layout (flow 4).

At the service layer, the GW collects information from the sensors to achieve the 3D layout of composite π-container, which further is exploited to provide important services (see Fig. 19.10).

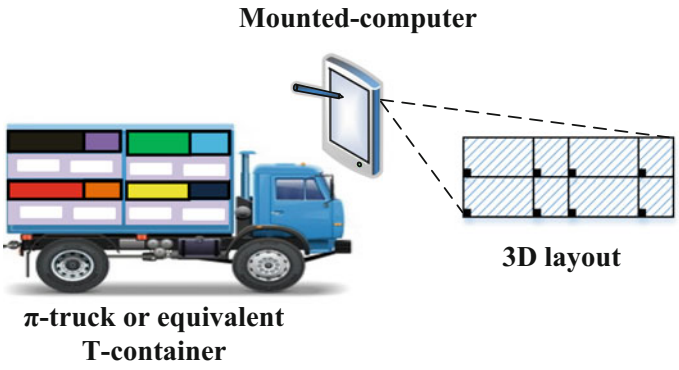


Fig. 19.10 IoT architecture for 3D layout retrieval at T-container level

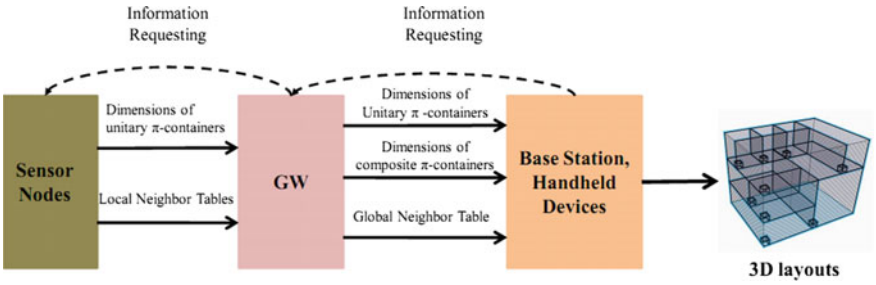


Fig. 19.11 Information exchange among sensors, GW and processing

19.4.2 An Information Flow Framework to Retrieve 3D Layouts

In this framework (see Fig. 19.11), the required information includes the dimensions of π-containers and the proximity information obtained by neighborhood relationship among sensor nodes. In order to validate the framework, a simulation-based method has been created and described in our previous research work [35].

With such simple methodology for retrieving the 3D layout of composite π-container, the value-added IoT services can be developed to support the logistics processes in the PI. In the next subsection, those services are introduced.

19.4.3 Value-Added Services Enabled by Retrieved 3D Layouts

The retried 3D layout provides the exact 3D view of spatial distribution of composed π-containers in their blocks. In addition, the locations of equipped sensor node are available. Thus, the sensing functionality offers distribution of environment conditions inside these blocks. In this way, the status of π-containers is monitored continuously to prevent issues in real time. Such as, the π-container equipped with adequate sensors can detect a problem (e.g., abnormal high temperature), check for detection integrity with nearby π-containers when pertinent, and send an alert message to their agents or directly to the PIMS. In another instance, proximity π-containers carrying cargoes incompatible with each other (e.g., chemical products that can contaminate each other or cause an explosion) can be detected and avoided when a π-container arrives at a π-hub, it sends to the PIMS the list of elements incompatible with its cargo.

19.5 Conclusion and Future Works

With rapid development in the emerging IoT technology, this chapter proposes a design framework of developing an SOA for Physical Internet using IoT, which is actually motivated and strongly demanded from the logistics managers as they need the service provision to efficiently manage and control the logistics processes. In this context, we describe key infrastructure and proposition design mostly based on the advanced ICT innovation to enable IoT for such logistics service providing 3D layout of composite π-containers is presented as a case study to highlight the practical usage and merit of our proposed framework.

PI is an innovative concept in logistics. Several researches and studies since 2011 have contributed to demonstrate and give the proof of concept. Looking to the future roadmap, the goal in 2020 is to realize interoperability between networks and ICT applications for logistics. Obviously, application of IoT into brings huge benefits in terms of economy, environment, and society. In other word, the IoT enables the PI achieve its global sustainability goal. However, deploying and realizing this revolution technology faces significant challenges mostly from the technical perspectives. For instance, since the physical facilities of the PI are equipped with smart devices (e.g., RFID, sensors) usually powered by batteries, saving their energy to prolong their operation lifetime is important. These IoT devices can be requested anytime to provide information for higher layer in the SOA to develop the services. Thus they may be active always. To reduce the power consumption the WSN nodes will be asleep (low-power mode) most of the time and it will only wake up to acquire the sensor data or to talk to another node. Several important issues such as standardization, network type, the quality of service, and logistics data protection are expected to provide a basis for further research on IoT-based logistics services.

References

1. Crainic TG, Gendreau M, Potvin J-Y (2009) Intelligent freight transportation systems: assessment and the contribution of operations research. Transp Res C Emerg Techno 17(6):541–557
2. Zhou L, Lou CX (2012) Intelligent cargo tracking system based on the internet of things. In: Proceedings of 15th international conference network-based informatics systems, Sept 2012, pp 489–493
3. Schumacher J, Rieder M, Gschweidl M, Masser P (2011) Intelligent cargo-using internet of things concepts to provide high interoperability for logistics systems. Springer, Berlin, Germany, pp 317–347
4. Forcolin M, Fracasso E, Tumanischvili F, Lupieri P (2011) EURIDICE-IoT applied to logistics using the intelligent cargo concept. In: Proceedings of 17th international conference concurrent enterprising, June 2011, pp 1–9
5. Cosimato S, Troisi O (2015) Green supply chain management: practices and tools for logistics competitiveness and sustainability. The DHL case study. TQM J 27(2):256–276
6. Dekker R, Bloemhof J, Mallidis I (2012) Operations research for green logistics-an overview of aspect, issues, contributions and challenges. Eur J Oper Res 219(3):671–679
7. Montreuil B (2011) Toward a physical internet: meeting the global logistics sustainability grand challenge. Logistics Res 3(2–3):71–87
8. Montreuil B, Meller RD, Ballot E (2012) Physical internet foundations. IAFC Proc 45(6):26–30
9. Ballot É, Montreuil B, Meller RD (2014) The physical internet: the network of logistics networks. La Documentation Française, France
10. Landschützer C, Ehrentraut F, Jodin D (2015) Containers for the physical internet: requirements and engineering design related to FMCG logistics. Logist Res 8(1):8
11. Meller RD, Lin Y-H, Ellis KP (2012) The impact of standardized metric physical Internet containers on the shipping volume of manufacturers. IAFC Proc 45(6):364–371
12. Rd M, Yh L, Kp E, Lm T (2012) Standardizing container sizes saves space in the trailer. Center Excellence Logistics Distribution University, Arkansas, Fayetteville, AR, USA, Technical Report 01
13. Venkatadri U, Krishna KS, Ülkü MA (2016) On physical internet logistics: modeling the impact of consolidation on transportation and inventory costs. IEEE Trans Autom Sci Eng 13(4):1517–1527
14. Ülkü MA (2012) Dare to care: shipment consolidation reduces not only costs, but also environmental damage. Int J Prod Econt 139(2):438–446
15. Sang K, Hong K-S, Kim KH, Lee C (2017) Shipment consolidation policy under uncertainty of customer order for sustainable supply chain management. Sustainability 9(9):1675
16. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
17. Sternberg H, Norrman A (2017) The physical internet-review, analysis and future research agenda. Int J Phys Distrib Logist Manage 47(8):736–762
18. Montreuil B, Ballot E, Tremblay W (2014) Modular structural design of physical internet containers. Prog Mater H Res 13
19. Montreuil B, Ballot E, Tremblay W (2015) Modular design of physical internet transport, handling and packaging containers. In: Proceedings of progress in material handling research. MHI, Tokyo, Japan, pp 1–13
20. Lin Y-H, Meller RD, Ellis KP, Thomas LM, Lombardi BJ (2014) A decomposition-based approach for the selection of standardized modular containers. Int J Prod Res 52(15):4600–4672
21. Meyer GG, Främling K, Holmström J (2009) Intelligent products: a survey. Comput Ind 60(3):137–148
22. Sallez Y, Montreuil B, Ballot E (2015) On the activeness of physical internet containers. Springer, Cham, Switzerland, pp 259–269
23. Montreuil B, Meller RD (2010) Toward a physical internet: the impact on logistics facilities and material handling systems design and innovation. In: Gue K (ed) Progress in material handling research. Material Handling Industry of America, p 23

24. Tretola G, Biggi D, Verdino V (2015) A common data model for the physical internet. In: Proceedings of the 2nd international physics international conference, Paris, France, 2015, pp 160–176
25. Lin J, Yu W, Zhang N, Yang X, Zhang H, Zhao W (2017) A survey on internet of things: architecture, enabling technologies, security and privacy, and applications. IEEE Internet Things J 4(5):1125–1142
26. Miorandi D, Sicari S, De Pellengrini F, Chlamtac I (2012) INTERNET of things: vision applications and research challenges. Ad Hoc Netw 10(7):1497–1516
27. Xu LD (2011) Enterprise systems: state-of-the-art and future trends. IEEE Trans Ind Informat 7(4), 630–640
28. Sun C (2012) Application of RFID technology for logistics on internet of things. AASRI Procedia 1:106–111
29. Vazquez JI, Almeida A, Doamo I, Laiseca X, Orduæa P (2009) Flexeo: an architecture for integrating wireless sensor networks into the internet of things. In Proceedings of 3rd symposium ubiquitous computing and ambient intelligence, pp 219–228
30. Flügel C, Gehrmann V (2009) Scientific workshop 4: intelligent objects for the internet of things: internet of things—application of sensor networks in logistics. In: Proceedings of European conference ambient intelligence, pp 16–26
31. Tran-Dang H, Krommenacker N, Charpentier P (2014) Localization algorithms based on hop counting for wireless nano-sensor networks. In: Proceedings of the international conference on indoor positioning and indoor navigation (IPIN), Oct 2014, pp 300–306
32. Kubler S, Derigent W, Främling K, Thomas A, Rondeau É (2015) Enhanced product lifecycle information management using 'communicating materials'. Comput Aided Des 59:192–200
33. Kubler S, Derigent W, Thomas A, Rondeau É (2014) Embedding data on 'communicating materials' from context-sensitive information analysis. J Intell Manuf 25(5):1053–1064
34. Guinard D, Trifa V, Karnouskos S, Spiess P, Savio D (2010) Interacting with the SOA-based internet of things: discovery, query, selection, and on-demand provisioning of web services. IEEE Trans Serv Comput 3(3):223–235
35. Tran-Dang H, Krommenacker N, Charpentier P (2017) Containers monitoring through the physical internet: a spatial 3D model based on wireless sensor networks. Int J Prod Res 55(9):2650–2663
36. El Ghomali K, Elkamoun N, Hou KM, Chen Y, Chanet J-P, Li J-J (2013) A new WPAN model for NS-3 simulator. In: Proceedings of new information communication science technolology sustainable development France-China international workshop (NICST), p 8
37. Tran-Dang H, Krommenacker N, Charpentier P (2015) Enhancing the functionality of physical internet containers by wireless sensor networks. In: Proceedings of 2nd international physical internet conference (IPIC), Paris, France, July 2015, pp 86–98

Index

0–9
1553B Interface, 89
3D layout, 242, 245, 246, 248, 253, 255–257, 276–279

A
Access delay, 162, 167–169, 177
Ad hoc On-Demand Distance Vector (AODV), 168
Aloha, 147, 153
Avionics system, 89

B
BasicCAN, 31, 39
Big data, 212
Binary exponential backoff, 173, 175
Bluetooth, 122, 131, 136, 181, 184, 189–191, 194, 199, 211, 217
Broadcast, 51, 66, 68, 79, 90, 91, 95, 121, 142, 143, 148, 150–152, 221, 225
Bus arbitration, 37
Bus Control Unit (BCU), 89

C
CAMMAC-802.11, 161–163, 171
CAN bus, 68
Carrier free direct sequence ultra-wideband technology, 198
Carrier Sense Multiple Access (CSMA), 5, 6, 12, 36, 37, 111, 124, 134, 147, 150, 151, 157, 158, 174, 175, 231
Carrier-sense Multiple Access Arbitration on Message Priority (CSMA/AMP), 5, 68
Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), 173, 175, 231, 233, 235
Channel capacity, 185, 186
Cloud computing, 212
Clustering, 143, 157, 218–221, 225, 226, 229, 233, 237
Code-Division Multiple Access (CDMA), 148, 156, 157
Controller Area Network (CAN), 8–11, 17, 24, 29, 31–41, 44–48, 51–59, 62, 65, 68–72, 124
Convergecast, 142, 143
Cooperative communication, 161, 218

D
Dependability, 41, 115
Determinism, 5, 7, 9, 12, 13, 15, 40
Device-to-Device communication (D2D), 211
Digital bus interface, 73
Digital time division multiplex data bus, 86
Distributed Control System (DCS), 3, 4, 31, 43–45, 47, 53, 54, 56, 59, 62
Distributed Coordinate Function (DCF), 170, 173–175
Distributed Interference-Aware Relay Selection (DIRS), 161, 170, 172
Dual fieldbus, 47
Dynamic Sensor-MAC (DSMAC), 153, 154

D.-S. Kim and H. Tran-Dang, *Industrial Sensors and Controls in Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-04927-0

E
Electrnoic Control Unit (ECU), 17, 22, 26–28
End-to-end delay, 162, 168, 169, 219, 230, 232, 233, 235–237
Energy-aware, 220, 228
Ethernet, 5–7, 10–15, 37, 44, 49, 117, 212
Event-triggered message, 17

F
Factory automation system, 115
Factory Information Protocol (FIP), 39, 121, 122
Fault tolerance, 45, 73, 79, 110, 120
FFD coordinator, 103
Fieldbus, 4, 7–11, 13, 14, 65, 72–74, 115–124
Field Programmable Gate Array (FPGA), 14, 75, 86, 89, 91–93
FlexRay, 17–29
Frequency-Division Multiple Access (FDMA), 148, 156, 157
FullCAN, 39
Full Function Device (FFD), 103, 104, 108, 109, 144, 145, 147

H
Handoff, 167
High-bandwidth communication, 17
High data rate, 17, 45, 54, 82, 86, 108, 184, 185, 191, 192, 198, 199, 201
High reliability, 41, 44, 73, 74, 82, 130, 139
High security, 107, 185
HyPer-1553TM, 81–83, 86

I
IEEE 802.11, 111, 117, 120, 123, 161, 162, 170, 174, 175, 178, 179, 188
IEEE 802.15.3, 185, 187, 191, 192
IEEE 802.15.4, 103, 117, 121, 123, 130–132, 134, 144, 148, 218, 219, 230, 231, 233–235, 246, 276
IEEE 802.15.4a MAC, 219
Industrial control system SCADA, 3
Industrial environment monitoring, 139
Industrial Internet of Things (IIoT), 207, 217
Industrial networks, 3, 4, 6, 12, 17, 119, 137
Industrial Wireless Local Area Network (Industrial WLAN), 120
Industrial Wireless Sensor Networks (IWSN), 127
Information and Communication Technology (ICT), 213, 243, 260, 263, 265, 279
Interference Immunity, 185
Internet of Things (IoT), 207, 217, 262
Inter-node Interference, 161, 170, 172
Interoperability, 8, 15, 73, 80, 86, 119, 122, 192, 203, 207, 210, 217, 272, 275, 279
ISA100.11, 118, 119, 128, 130, 134–136, 211

L
Logistics, 111, 203, 207, 213, 241–245, 247, 256, 259–262, 264, 265, 267–269, 271–274, 276, 278, 279
Low cost, 15, 44, 73, 108, 127, 131, 138, 173, 181, 186, 187, 190, 192, 199, 202, 210, 245, 259, 272
Low power, 74, 75, 130, 131, 147, 181, 184, 185, 190, 192, 197–201, 211, 220, 234
Low Rate-Wireless Personal Area Network (LR-WPAN), 120, 121

M
Machine-to-Machine communication (M2M), 211
Manchester bi-phase encoder/decoder, 78, 86
Manchester II bi-phase, 77, 90, 94
Master-slave query-response cycle, 67
Medium Access Control (MAC), 5, 36, 101, 110, 119, 132, 141, 146, 173, 174, 178
Mesh topology, 105, 145, 146
Message priority, 5, 68
Message scheduling, 20, 22
MIL-STD-1553 protocol, 78, 82
Mobile-robot, 137
Modbus, 9, 10, 13, 14, 44–51, 53, 54, 58, 61, 62, 65–68, 70–72
Multi-band OFDM, 187, 198
Multi-Band OFDM Ultra-Wideband Technology (MBOFDM), 198
Multi-cast, 221
Multi-channel Access, 161, 163
Multi-hop, 105, 134, 145, 150, 156, 164, 165, 218, 221, 231, 234, 244, 247
Multi-master serial bus, 41
Multipath fading, 198

N
Narrow band, 111, 183–186, 188, 190, 198, 203
Network topology, 18, 33, 81, 104, 109, 117, 143, 144

P
Physical Internet, 213, 241, 242, 244, 256, 257, 260, 261, 263, 265, 266, 267, 279
Plant floor, 136, 215

Point Coordinate Function (PCF), 173, 174, 176
Protocol conversion interface, 65, 70–72
Pulse Modulation, 199

R
Radio Frequency Identification (RFID), 243, 244, 263, 273, 279
Real time, 7, 43, 44, 74, 82, 85, 120, 136, 138, 139, 191, 208, 213, 214, 243, 260, 265, 268, 269, 279
Real-time communication, 49, 115
Real-time routing, 220, 228, 242
Reduced Function Device (RFD), 103, 104, 109, 144, 147
Redundancy, 17, 18, 20, 27, 35, 38, 44–46, 48, 49, 53–55, 62
RFID, 111, 207, 210, 211, 213
R-Fieldbus, 119, 120
Robustness, 11, 38, 73, 107, 108, 111, 120, 182, 189, 197, 199, 200

S
Sensor-MAC, 149, 155
Service-Oriented Architecture (SOA), 262
Ship engine system, 43, 62
Short-duration pulse, 182, 197
Sift, 153
Smart city, 214
Smart containers, 242, 245
Smart grid, 215
Spatial diversity, 161, 162
Star topology, 104, 144
Superframe, 174
Supply chain, 138, 207, 213, 241–243, 260, 261, 264

T
Throughput, 3, 37, 80, 83, 102, 106, 123, 143, 154, 157, 162–166, 170, 171, 177, 182, 186, 197, 198, 218
Time-Division Multiple Access (TDMA), 17
Time-triggered message, 17, 41
Traffic-Adaptive MAC (TRAMA), 152
Tree topology, 104, 105, 117, 145
Turbo-1553, 83, 84, 86

U
Ultra-Wideband (UWB), 181, 182, 190, 191, 197, 199, 201, 202

W
WiFi, 111, 112, 140, 199, 221
WiMedia UWB, 192
Wireless Ad hoc networks, 275
Wireless fieldbus, 115–120, 122–124
WirelessHART, 118, 119, 128, 130, 132–134, 136, 211, 218, 230
Wireless Personal Area Network (WPAN), 181
Wireless Sensor Network (WSN), 101, 103, 108, 109, 111, 124, 127, 137–140, 142, 143, 157, 207, 211, 217, 232, 244, 247, 256, 263
WiseMAC, 150, 151

Z
ZigBee, 108, 111, 112, 119, 122, 128, 130–132, 136, 211, 218