IntechOpen

Wireless Sensor Networks

Technology and Applications

Edited by Mohammad Matin

WIRELESS SENSOR NETWORKS – TECHNOLOGY AND APPLICATIONS

Edited by **Mohammad A. Matin**

Wireless Sensor Networks - Technology and Applications
http://dx.doi.org/10.5772/1100
Edited by Mohammad Matin

Contributors

Hamida Seba, Besem Abid, Souleymane M'Bengue, Kia Makki, Ali Bakhtiar, Jie Wang, Qinghua Gao, Yan Yu, Hongyu Wang, Minglu Jin, Yuanyuan Zeng, Kai Xiang, Deshi Li, Charalambos Sergiou, Vasos Vassiliou, Faisal Ahmed Al-Nasser, Magdi Mahmoud, Hussein Kdouh, Gheorghe Zaharia, Christian Brousseau, Guy Grunfelder, Ghaïs El Zein, Hanna Farhat, Fabrice Theoleyre, Benoit Darties, Nazim Abdeddaim, Luca Bencini, Stefano Maddio, Letícia Maria Bolzani Poehls, Alex Pinto, Fabian Luis Vargas, Carlos Montez, Sofiane Moad, Nizar Bouaballah, Yassine Hadjadj, Bruno Sericola, Qiang Zhou, Paulo Sausen, Airam Sausen, Fabiano Salvadori, Mauricio De Campos, Renê Emmel Júnior, Mario Collotta, Jeisa Domingues, Ehsan Sheybani

Notice

First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Wireless Sensor Networks - Technology and Applications
Edited by Mohammad Matin

p. cm.

ISBN 978-953-51-0676-0

eBook (PDF) ISBN 978-953-51-5573-7

Meet the editor

Dr Mohammad A. Matin is currently working at EEE department, Institut Teknologi Brunei (ITB) as an Associate Professor. He obtained his MSc degree in digital communication from Loughborough University, UK and PhD degree in wireless communication from Newcastle University, UK. He has published over 50 refereed journals and conference papers. He is the author of three academic books and five book chapters. He is currently serving as a member of editorial board of several international journals such as IET Wireless Sensor Systems (IET-WSS), IJCTE, JECE, etc. He has received a number of Prizes and Scholarships including the Best student prize, Commonwealth Scholarship and Overseas Research Scholarship. His current research interests include UWB communication, wireless sensor networks, cognitive radio, EM modeling, and antenna engineering.

Contents

Preface

A Wireless Sensor Network (WSN) represents the next technological revolution which distinguishes from other wireless or wired networks through its capability of interaction with the environment. This network has been proposed for various applications including monitoring the environment, home and business smart environments, better management of cities in areas like traffic control, intelligent transportation, search and rescue, disaster relief, and localization systems. These applications require a large amount of battery-powered wireless sensors, and are usually designed for long-term deployments with no human intervention. Subsequently, energy efficiency is one of the main design objectives for these sensor networks. This book offers the basics as well as advanced research materials for wireless sensor networks.

Chapter 1 deals with the proper node placement in the monitored field to ensure good sensing coverage and communication connectivity. It presents and analyzes four specific ways that nodes can be placed on a grid and compares the performance of congestion control algorithms under these placements.

Chapter 2 attempts various possible approaches that designers might adopt in order to provide energy efficient WSN. Some approaches offer to define a trade-off between opposing aims, such as QoS and energy consumption. Each approach presented in this Chapter is aiming at solving at least one identified cause of extensive energy consumption. The approaches are using the WSN's different design levels to increase the network's lifetime, QoS, and/or optimize other points that may or may not suit the designer.

Chapter 3 examines a radio-triggered wake-up circuit and explores its application in the power management of the WSNs. It should be mentioned that the radio-triggered wake-up circuit can be used in other applications, such as the synchronization of the WSNs.

Chapter 4 designs some MAC solutions to deal with real time applications in WSNs where mobile nodes are implemented in the network. WSN MAC protocols are application dependent and focus on energy consumption and transmission latency.

Chapter 5 explains how the cluster-tree structure impacts the network capacity and the energy Consumption. This chapter also presents a localized algorithm to find autonomously accurate parameters values, to operate always around the optimal situation.

Chapter 6 deals with the problem of reliability modelling with wireless sensor network (WSN), which is rapidly becoming a platform for applications including antiterrorism, smart spaces, numerous military sensing and command and control applications, and entertainment. Inherent in these safety-critical applications is a priority and urgency of the information or messages. To the best of the authors' knowledge, there is no systematical research done so far to unify energy consume and message delay into reliability modelling for WSN. The work in this chapter differs from the previous work in that it proposes a model of the system and an integrated model of the task which consider energy consume and message delay for the safety-critical application, introduces both the energy factor function and time factor function, and also establishes an integrated reliability model of WSN based on a task. The illustration of modelling suggests that the method studied has a directive influence to both task division and topology selection of WSN system.

Chapter 7 introduces a framework for integrating WSNs and the Internet at service level, by allowing the interoperability of their services.

Chapter 8 presents data reduction process in WSN which is a challenging task as data exists in the form of continuous stream (infinitely large data set) where the adaptation and prediction has to be performed online i.e. at a given instance of time not all the information is available for processing. Typically, the spatial and temporal relations among the data sources in WSN are exploited to achieve fair data reduction rates. In essence, it can be said that data reduction is one of the most effective ways to conserve energy in Wireless Sensor Networks.

Chapter 9 presents computationally low power, low bandwidth, and low cost filters that will remove the noise and compress the data so that a decision can be made at the node level.

Chapter 10 presents clustering scheme which is based on selecting the node that reduces the packet size among all the active nodes in the system. The sink selects the node which minimizes the total amount of data as a cluster head (CH), therefore increasing the efficiency of the compression technique by sending only the difference, rather than the complete data value, to the CH.

Chapter 11 focuses on some of the main concerns for design in mission-critical monitoring application scenarios, and brings forward efficient methods and solutions according to network connectivity, dynamic application scenarios. Major research challenges and open research issues in mission-critical monitoring applications of WSNs are also outlined.

Chapter 12 outlines the design, optimization and development of a practical solution for application to the Cultural Heritage monitoring and control. The overall system was addressed in terms of the experiences platform, network issues related both to the node's communication protocol and gateway operations up to the remote user's suitable interface. In particular, the presented solution is installed in several museums and it is used to monitor the art objects during their transport from a museum to another. The experimental results highlight a noticeable performance as far as the data collecting reliability, the system robustness and the usability are involved.

Chapter 13 presents the application of wireless technologies to the shipboard monitoring system. A measurement campaign has been carried out on board a ferry to determine path loss models. Based on the measurement results and the particularities of the environment, a hierarchical zone-based architecture has been proposed for a large shipboard WSN.

Chapter 14 presents a monitoring system that is applied to underground power electrical substation. The system presented in this chapter allows choosing the desired communication transmission mode: wired, wireless, or both. Performance results show that our system could well be applied for monitoring and fault detection in electrical underground network grid systems.

Chapter 15 presents some solutions to certain road monitoring problems unresolved by several works presented in literature, in which streets monitoring are realized using cameras only. Cameras can produce several problems when weather conditions are not optimal (poor visibility, fog or heavy rain). Road monitoring and WSNs have been extensively explored in their respective domains, however, the combination of them has not been sufficiently studied until recent years. The challenge of this chapter is to advance the state of the art introducing an innovative approach that allows monitoring real-time road traffic and, at the same time, dynamically manages network topology and power consumption through a fuzzy based algorithm.

Chapter 16 introduces some control strategies and concepts from both traffic and control engineering point of view and provides some simulations to support the theoretical discussion.

I hope that this book will serve as a comprehensive reference for graduate students as well as senior undergraduate students and that it will be useful as a learning tool for research in this exciting field.

Mohamamd A. Matin
Institut Teknologi Brunei,
Brunei Darussalam

Section 1

Design and Analysis for Deployment

Chapter 1

Efficient Node Placement for Congestion Control in Wireless Sensor Networks

Charalambos Sergiou and Vasos Vassiliou

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48201

1. Introduction

Wireless sensor nodes are small, embedded computing devices that interface with sensors/ actuators and communicate using short-range wireless transmitters. Such nodes act autonomously, but cooperatively to form a logical network, in which data packets are routed hop-by-hop towards management nodes, typically called sinks or base stations. A Wireless Sensor Network (WSN) comprises of a potentially large set of nodes that may be distributed over a wide geographical area, indoor or outdoor. Wireless Sensor Networks (WSNs) enable numerous sensing and monitoring services in areas of vital importance such as efficient industry production, safety and security at home as well as in traffic and environmental monitoring. Traffic patterns in WSNs can be derived from the physical processes that they sense. WSNs typically operate under light load and suddenly become active in response to a detected or monitored event. Early research studies in WSNs targeted military applications, especially for battlefield monitoring. In the last few years, due to the progress of low powered units and improvements in radio technologies, wireless sensor networks technologies have gained momentum. WSNs are now being deployed in civilian areas and being used for habitat observation ([1], [2]), health monitoring ([3]), object tracking ([4], [5]) etc. In addition, lately, there is an emergence of mission-critical applications ([6]).

Emergence of mission-critical and information demanding applications in WSNs renders performance control essential, for mission accomplishment. Heavy traffic load is a major factor that affects the performance of any type of network. The situation worsens in low-powered, unreliable WSNs. Thus, a prominent factor that under specific circumstances, can improve or deteriorate the performance of WSNs, can be the way that nodes are placed on the monitored field. Proper node placement is essential to ensure good sensing coverage and communication connectivity.

In this work we present and analyze four general ways that nodes can be placed on a grid and we compare the performance of a representative number of routing and congestion control algorithms under these placements.

The examined algorithms are ESRT ([7]), Sen- TCP ([8]), Directed Diffusion ([9]), HTAP ([10]) and Flooding. These algorithms are evaluated under four different placements. Simple Diffusion, Random, Grid and Biased- Random. Algorithms and placements are described in detail, in the next sections.

Initial results of this work have been presented in ([11] and [12]). In this paper we extent the number of evaluated algorithms in order to present a complete and solid work. Thus, we include an algorithm that represents the category of "reliable data transport" (ESRT), as well as a generic routing algorithm ("flooding"). Hence, in this paper, we provide representative results from all the categories of congestion control and reliable data transport algorithms in WSNs, under different placements.

Simulation results show that the performance of specific algorithms can be improved under specific placements. In particular, algorithms that employ multiple or alternative paths for forwarding the excess traffic from source to sink are favored by specific placements.

2. Related work

Several node placements have been proposed in literature concerning WSNs.

Younis et al ([13]) present a survey for strategies and techniques for node placements in WSNs and provide a categorization of the placement strategies into static and dynamic, depending on whether the optimization is performed at the time of deployment or while the network is operational.

Toumpis et al ([14]) provide an optimal deployment of large wireless sensor networks so as to minimize the number of nodes that is needed in order to transmit data from multiple sources to multiple sinks.

In ([15]) authors evaluate the tolerance against both random failure and battery exhaustion from the viewpoint of stochastic node placement. They consider three typical types of stochastic sensor placement: Simple diffusion, Constant Placement and R- Random placement.

In ([16]) authors studied the problem of determining the critical node density for maintaining k-coverage of a given square region. They have considered three different deployment strategies: Poisson point process, uniform random distribution and grid deployment and have shown that the two random strategies have identical density requirements for k-coverage. They also showed that grid deployment requires less node density than the two random deployments strategies in order to achieve the same level of coverage degree.

In ([11]) authors present a performance study for congestion control between three different algorithms under different node placements. Algorithms that employ three different techniques for congestion mitigation in WSNs. "Traffic control", "resource control" and multipath routing. Simulation results show that the performance of "resource control" algorithms is affected by different node placements.

In ([12]) authors study the energy utilization performance of HTAP algorithm([10]) under specific node placements, in correlation with Directed Diffusion ([9]) algorithm. Simulations results show that the performance of HTAP, a "resource control" algorithm, is improved when nodes are densely deployed near hotspots.

3. Congestion control in WSNs

Congestion in WSNs occurs when the offered load is temporarily higher than the load which node(s) resources can process in a certain amount of time.

Congestion in WSNs can be categorized in two types. The first type of congestion happens in the medium. In this type, two or more nodes attempt to transmit simultaneously and as a result collisions of packets occur in the medium. This type of congestion is normally faced through enhancements in the MAC layer (e.g phase shifting that appeared in ([17]).

The other type of congestion happens when the queue or the buffer of a node used to hold packets to be transmitted, overflows. In such case packets drops happen, which is a highly undesirable situation in low powered WSNs. Solutions for this type of congestion lie in upper layers, like network or transport layer.

Generally, congestion in WSNs is mitigated by three categories of algorithms. "Traffic Control", "resource control" and "reliable data transport". "Traffic control" algorithms, affect the data rate of source nodes in order to reduce the traffic in the network when congestion occurs. Algorithms that employ this method, attempt, normally usually backpressure messages, to inform sources to reduce their data rate, in order to absorb the already high load and avoid packet drops.

On the other hand, "resource control" algorithms employ a different method in order to mitigate congestion. In this case, these algorithms attempt to take advantage of the already dense placement of WSNs, as well as the plethora of nodes that are in sleep state, by creating alternative or multiple paths to the sinks, in order to route the excess data. This type of algorithms do not affect the rate with which sources inject traffic in the network.

Finally, a different category is "reliable data transport" algorithms. This type of algorithms, typically run on the transport layer and focus on reliability. Although they are not "pure" congestion control algorithms, they can be considered as so, since congestion is a condition that affects significantly the reliability of WSNs.

Besides these categories there is another type of algorithms that attempts to create multiple paths in order to ease the transportation of data from source to sink. Although algorithms that fall in this category, cannot be considered as congestion control algorithms, we study them, since multiple paths can assist the network to balance the load and avoid congestion occurrence.

Thus, in this work we study the behavior of a representative algorithm of each category when nodes are placed under different placements. Specifically we employ SenTCP ([8])

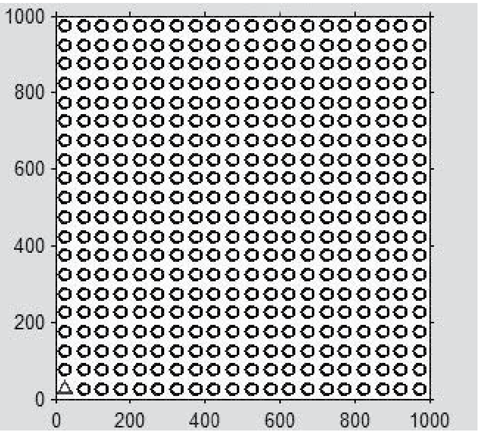


Figure 1. Grid Placement

as a "traffic control" algorithm, HTAP ([10]) as a "resource control", ESRT [7]) as a "reliable data transport" algorithm as well as "Directed Diffusion" ([9]) as a "multiple path creation" algorithm. Furthermore, we also study "flooding algorithm" which is a generic routing algorithm, for comparison purposes.

4. Node placements analysis

It is clear that node density is only one factor that affects network topology. The actual placement of nodes is also significant, as it is shown in ([13] and [18]). The placement of nodes affects the ability of a network to correctly sense an event while it also affects the number of possible disjoint paths towards the sink(s).

Thus, we claim that the placement of sensor nodes on a monitored field, is a factor that it is possible to affect the overall performance of the network.

Placement of nodes in a network can be divided into three major categories concerning the way that nodes are placed in the field. These are the deterministic node placement, the semi- deterministic node placement and the non- deterministic (stochastic) node placement. In this work we choose to place nodes in four different placements in order to cover all categories. A deterministic placement (Grid), a semi- deterministic (Biased Random) and two non- deterministic (Simple Diffusion and Random).

4.1. Deterministic node placement

In deterministic node placement, nodes are placed on exact, pre- defined points on a grid or in specific parts of the grid. Usually, deterministic or controlled node placement is specified by the type of nodes, the environment in which the nodes will deploy, and the application. Therefore, in applications like Sensor Indoor Surveillance Systems or Building Monitoring, nodes must be placed manually ([13]) (either by hand or by robots). In this work we employ Grid Placement as appears in Fig. 1. In this placement nodes are placed strictly on the lines of a grid.

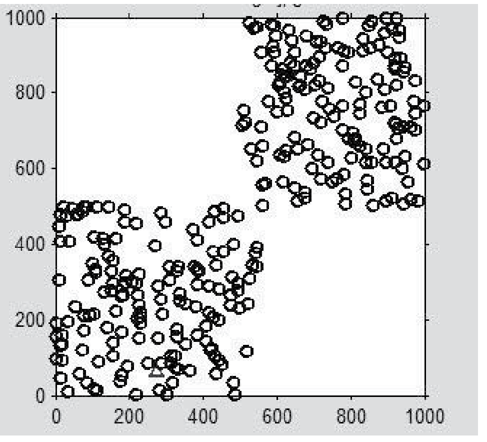


Figure 2. Biased Random Placement

4.2. Semi- deterministic node placement

Semi- deterministic placement is the placement, where, although individual nodes are placed in a non- deterministic way on the grid (e.g random) the areas where nodes are going to be spread are deterministic. This means that in a microscopic way the placement of nodes is non- deterministic while in a macroscopic way the placement is deterministic. In this paper we employ biased- random placement, where nodes are placed in two specific areas (near source and near sink). Note that the actual node placement is performed in random way in these areas (Fig. 2)

4.3. Non- deterministic node placement

Deterministic placement is not so realistic when many sensor nodes are placed in a large area. In such a situation, stochastic placement is needed. In this paper we employ two stochastic placements. Simple Diffusion and Random placement

Simple Diffusion: This node placement emulates the distribution of nodes when they are scattered from air e.g from airplane (Fig. 3). Simple diffusion is analytically explained in ([15]).

Random Placement: This is a commonly used topology and sensor nodes are placed so that their density is uniform. (Fig. 4)

5. Congestion control algorithms

In different studies ([17], [19]) it is observed that the number of nodes with occupied queues grows, if congestion gets worse.

Several transport control schemes and algorithms have been proposed in the literature ([7, 8, 10, 17, 19–25]). Their objectives and approach differs. The vast majority of them ([7, 8] and [17] to [22]) react to congestion with rate limiting techniques ("traffic control" algorithms). Others deal with the problem by increasing the resources ([10], [26], [27]) ("resource control"

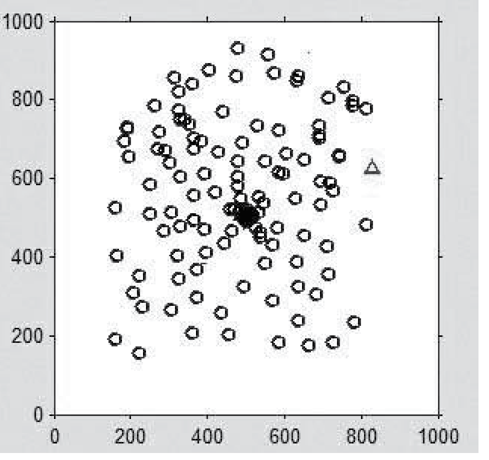


Figure 3. Simple Diffusion Placement

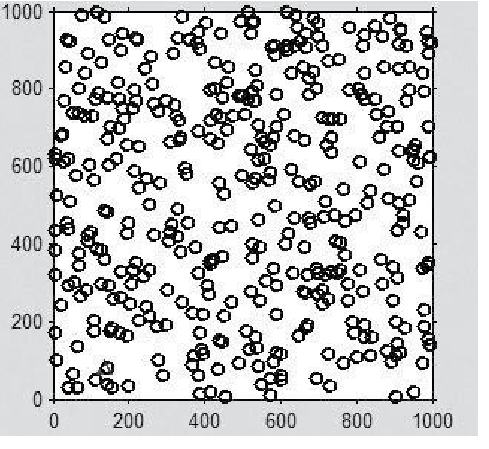


Figure 4. Random Placement

algorithms). Some focus more on reliability issues instead of congestion ([7, 24, 25]) ("reliable data transport" algorithms). In our evaluation we consider one algorithm from each major category: ESRT ([7]) which is "traffic control" algorithm focusing on reliability, SenTCP ([8]) which is a "traffic control" algorithm focusing on congestion, and HTAP ([10]) which is "resource control" algorithm. We also employ Directed Diffusion ([9]) algorithm, an algorithm not explicitly designed for congestion control but it can be considered as so, since it employs a combination of "traffic control" and "resource control" techniques in order to provide multiple disjoint paths to the sink. Finally we also evaluate the performance of a generic routing algorithm, the "plain flooding".

Short description of employed algorithms follows.

5.1. Sen- TCP (Sensor Based TCP)

SenTCP is an open-loop hop-by-hop congestion control protocol with two special features:

1. It jointly uses average local packet service time and average local packet inter-arrival time in order to estimate current local congestion degree in each intermediate sensor node. The use of packet arrival time and service time, not only precisely calculates congestion degree, but effectively helps to differentiate the reason of packet loss occurrence in wireless environments, since arrival time (or service time) may become small (or large) if congestion occurs.
2. It uses hop-by-hop congestion control. In SenTCP, each intermediate sensor node will issue feedback signal backwards and hop-by-hop. The feedback signal, which carries local congestion degree and the buffer occupancy ratio, is used for the neighboring sensor nodes to adjust their sending rate in the transport layer. The use of hop-by-hop feedback control can remove congestion quickly and reduce packet dropping, which in turn conserves energy.

5.2. ESRT (Event to Sink Reliable transport)

ESRT aims at providing reliability from sensors to sink while supporting congestion control simultaneously. It is an end-to-end algorithm trying to guarantee a desired reliability through regulation of sensors reporting frequency. It provides reliability for applications and not for each single packet. The sink uses congestion feedback from sensor nodes to broadcast a notification to adjust the reporting rate with two goals: i) to receive a sufficient number of nodes from the sink, and ii) to receive only as many packets as necessary in order to avoid congestion and save energy. ESRT runs on the sink, with minimal functionality required at resource constrained sensor nodes. ESRT protocol operation is determined by the current network state, based on the reliability achieved and congestion condition in the network. Firstly, it needs to periodically compute the factual reliability r according to successfully received packets in a time interval. Secondly, ESRT deduces the required sensor report frequency f from r. Finally, ESRT communicates f to all sensors through an assumed channel with high power. ESRT identifies 5 characteristic regions:

- No Congestion, Low reliability
- No Congestion, High reliability
- Congestion, High Reliability
- Congestion, Low Reliability
- Optimal Operating Region (OOR) which essentially translates to No Congestion, Medium-High Reliability

The target is to identify network's current state and bring it in OOR (Optimal Operating Region). If the event-to-sink reliability is lower than the required, ESRT adjusts the reporting frequency of source nodes aggressively in order to reach the target reliability level as soon as possible. If the reliability is higher than required, then ESRT reduces the reporting frequency conservatively in order to conserve energy while still maintaining reliability. This self-configuring nature of ESRT makes it robust to random, dynamic topology of WSNs. An additional benefit resulted from ESRT is energy-conservation since it can control the sensor reporting frequency. ESRT presents some disadvantages:

1. ESRT regulates report frequency of all sensors using the same value. It may be more reasonable to use different values, since each sensor node may have different contribution to congestion.
2. It assumes and uses a channel (one-hop) with high power that will influence the on-going data transmission.
3. It mainly considers reliability and energy-conservation. Feedback latency dependents on the network's size and may not scale in very large sensor networks.

5.3. Directed diffusion

Directed Diffusion is a data centric protocol because all communication is for named data. All nodes in a directed diffusion-based network are application-aware. This enables diffusion to achieve energy savings by selecting empirically good paths (small delay) by caching and processing data in-network (e.g., data aggregation). Directed diffusion consists of four (4) basic elements:

- interests
- data messages
- gradients
- reinforcements

An interest message is a query from a sink node to the network, which indicates application demands. It carries a description of a sensing task that is supported by a sensor network. Data in sensor networks is the collected or processed information of an event (e.g. physical phenomenon), it is named (addressed) using attribute-value pairs, while a sensing task is diffused throughout the sensor network as an interest for named data. This dissemination sets up gradients within the network designed to "draw" events (i.e., data matching the interest). A gradient is direction state, created in each node that receives an interest. This direction is set toward the neighboring node from which the interest was received. Events start flowing towards the sinks of interests along multiple gradient paths. To improve performance and reliability, the empirically "good paths" (e.g small delay) are reinforced by the sink and their data rate increases. On the other hand unreliable paths (e.g high delay) are negatively reinforced and pruned off.

5.4. HTAP (Hierarchical Tree Alternative Path)

HTAP is scalable and distributed framework for minimizing congestion and assuring reliable data transmissions in event based WSNs. As such it does not employ rate limiting actions, but tries to maintain a high level of packets data rate while minimizing packet losses. It is based on the creation of alternative paths from the source to sink ("resource control"), using the plethora of network's unused nodes, in order to safely transmit the observed data. The creation of alternative paths involves several nodes which are not in the initial shortest path from the source to the sink. The use of these nodes leads to a balanced energy consumption, avoiding the creation of "holes" in the network and prolonging network lifetime. The HTAP algorithm consists of four major parts.

- Flooding with level discovery functionality: Through this procedure, each node discovers its neighbor nodes and updates its neighbor table. In addition, sensor nodes are placed in levels from the source to the sink.
- Alternative Path Creation Algorithm: In order to avoid congestion, each candidate congested receiver sends a backpressure packet to the sender. Thus, the sender stops the transmission of packets to the candidate congested receiver and searches in its neighbor table to find the least congested receiver, in order to continue the transmission of data. The dynamic change of the receivers, leads to the creation of new routes from the source to the sink.
- The Hierarchical Tree Algorithm: A hierarchical tree is created, beginning at the source node. Connection is established between each transmitter and receiver using a 2-way handshake. Through this packet exchange, the congestion state of each receiver is communicated to the transmitter.
- Handling of Powerless (Dead Nodes): Special care is taken in HTAP algorithm concerning the nodes that their power is getting exhausted. Thus, when a node is going to exhaust its power, it is immediately extracted from the network and the tables of its neighbor nodes are updated.

6. Performance evaluation

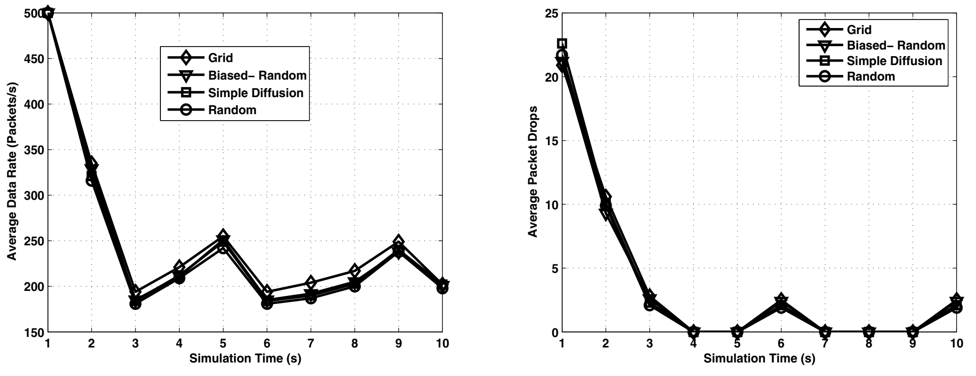
To evaluate the selected algorithms under the proposed topologies, a series of simulations using ns-2 [28] simulator, has been conducted.

6.1. Simulation environment

In all scenarios we choose to deploy nodes within a square area of size 1000m x 1000m. The results presented are the average of 20 runs for each measurement point. In each set of runs, the parameters of Table 1 were kept stable while increasing the number of nodes in the topology to make a dense network with strong connectivity.

X distance (m)	1000
Y distance (m)	1000
Transmit Power (mW)	600
Receive Power (mW)	600
Sensitivity Threshold (dBm)	-81
Path Loss Coefficient	3.5
Node CPU (MHz)	4
Radio Freq. (MHz)	433
Data packet	128 bytes
Control Packet	50 bytes
MAC layer	CSMA/CA
Initial Node Energy	1 Joule

Table 1. Simulation Parameters



(a) SenTCP:Average Data Rate (packets/s)

(b) SenTCP:Average Packet Drops

Figure 5. SenTCP:Average Data Rate (packets/s) and Average Packet Drops

The evaluation of node placements has been performed by studying four basic QoS parameters. These are: Average Packet Loss, Mean Packet's Delivery Delay, Average Data Rate, as well as the Percentage of Network's Remaining Energy after the point where the network, due to routing "holes" is not able to forward a single packets from source to sink.

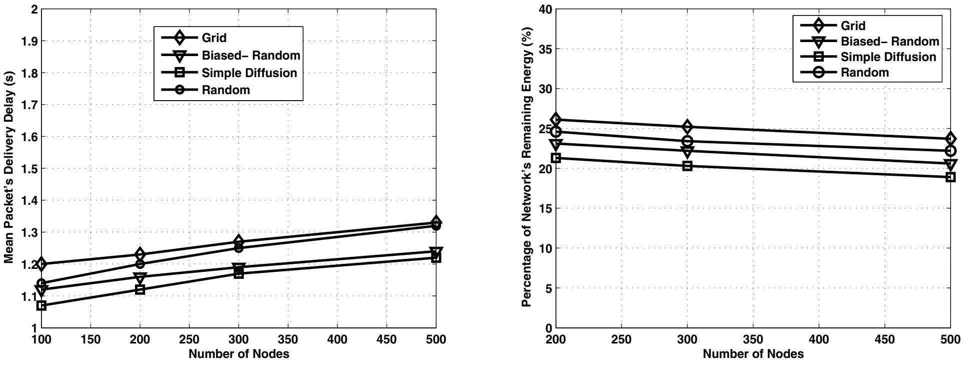
6.2. SenTCP evaluation

"Traffic control" is a method used in different algorithms to alleviate congestion in WSNs. SenTCP is one of these algorithms. In Fig. 5a we record the average data rate in all four examined topologies under SenTCP algorithm.

Initially, when 500 packets/s are injected in the overloaded network, the network experiences a heavy load situation and data rate falls rapidly in order to control the situation. SenTCP then slowly increases the rate until the occurrence of a new packet drop. It is clear that in all four topologies the network exhibits similar attitude and performance. This indicates that this parameter is slightly affected by nodes' placement. This is true, taking into account that SenTCP employs average local packet inter-arrival and packet service time, to detect congestion and "traffic contol" method to mitigate it.

Next we study packet drops. Packet drops is one of the most significant events in terms of performance control and their occurrence indicate a problem in the network. In Fig. 5b we record Average Packet Drops vs Simulation Time. According to this figure, the attitude of SenTCP algorithm is not affected by different placements. This result is an indication that "resource control" algorithms are not affected by different placements concerning their congestion mitigation ability.

Also a significant parameter concerning performance control, is the minimization or reduction of delays in the network. In Fig. 6a we present the mean time for the transmission of packets from source to sink. It is obvious that as the hop number increases, mean time increases too, since hop count is bigger. Considering that, algorithms like SenTCP use the shortest path to transmit their data, it is expected that the placement which is able to provide the shortest path



(a) SenTCP:Mean Packet's Delivery Delay (b) SenTCP:Percentage of Network's Remaining Energy

Figure 6. SenTCP:Mean Packet's Delivery Delay and Percentage of Network's Remaining Energy

will exhibit the best performance. This happens in Simple Diffusion placement followed by Biased- Random.

The last parameter that we investigate is the percentage of network's remaining energy at the moment where the network is not able to transmit a single packet from source to sink (network stalls). This metric is an indication of whether the network has managed to uniformly utilize its resources, avoiding the creation of network connectivity "holes" (Fig. 6b).

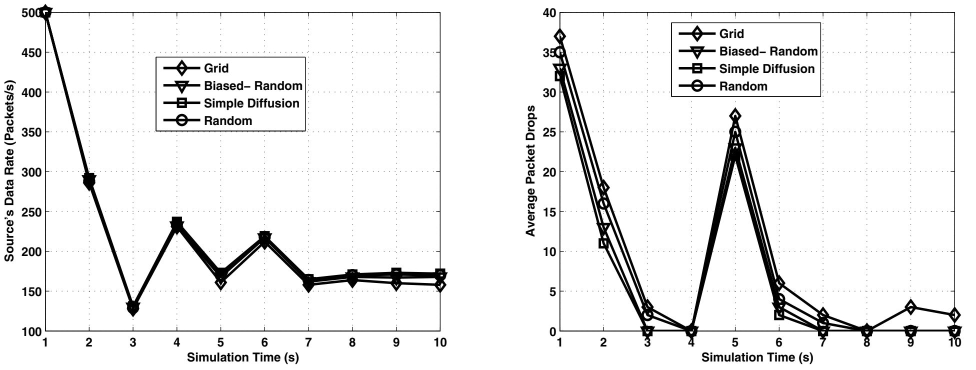
As it is presented in this graph, the network utilizes most of its energy in Simple Diffusion placement while it consumes the least energy in Grid placement. This proves that in Simple Diffusion placement where the nodes are scattered around the sink, the network is able to utilize more uniformly its resources by finding more available paths from source to sink compared to Grid Placement.

6.3. ESRT evaluation

ESRT is an algorithm that focuses on reliability. It is an end-to-end algorithm that runs on the sink and in case of congestion regulates report frequency (data rate) of all sensors using the same value. Fig. 7a presents the average data rate.

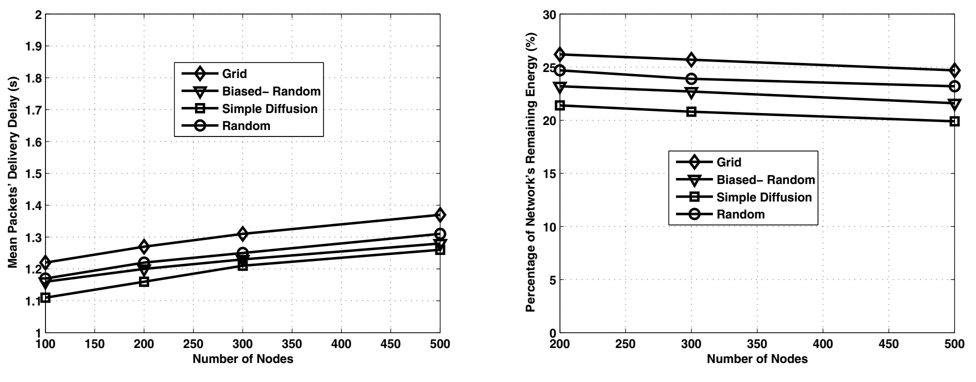
Since ESRT is an algorithm that throttles data rate in order to mitigate congestion, it is expected that average data rate is slightly affected by node placements. The same happens with packet drops (Fig. 7b).

On the other hand mean packet's delivery delay" is a parameter that its attitude could be related with nodes placement. Efficient nodes placement can reduce the mean time for the transmission of packets. As it is presented in Fig. 8a, Simple Diffusion placement and Biased-Random placement (as with SenTCP) are the placements that provide the shortest paths from source to sink and normally present the least delay. Contrary, Grid Placement is a placement that provides longer paths and this is the reason that the delay in this placement is the biggest.



(a) ESRT: Average Data Rate (packets/s) (b) ESRT:Average Packet Drops

Figure 7. ESRT: Average Data Rate (packets/s) and Average Packet Drops



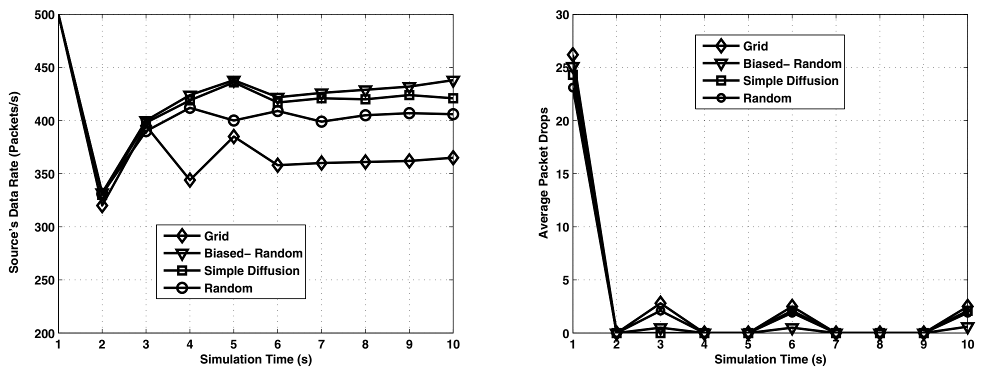
(a) ESRT:Mean Packet's Delivery Delay (b) ESRT:Percentage of Network's Remaining Energy

Figure 8. ESRT:Mean Packet's Delivery Delay and Percentage of Network's Remaining Energy

Concerning the percentage of network's remaining energy (Fig.8b) we notice that ESRT presents the same attitude as with SenTCP. ESRT when runs on Simple Diffusion and Biased- Random placements presents better performance in comparison with the other two placements. The reason is the same as with SenTCP. These placements provide a bigger number of disjoint paths from source to sink, that, when the nodes that form the initial paths are totally power exhausted, the network is still able to find other paths to forward data to sink.

6.4. Directed Diffusion evaluation

Directed Diffusion is an algorithm that mitigates congestion in an indirect way. Initially, it sends an upstream data message to multiple nodes, forming multiple paths and then, with reinforcement and negative reinforcement attempts to reduce the number of paths, to a small number, based on their empirically observed performance. Through this reduction of paths, it controls the data rate of the paths and consequently the network's data rate.



(a) Directed Diffusion: Average Data Rate (packets/s)

(b) Directed Diffusion:Average Packet Drops

Figure 9. Directed Diffusion:Average Data Rate (packets/s) and Average Packet Drops

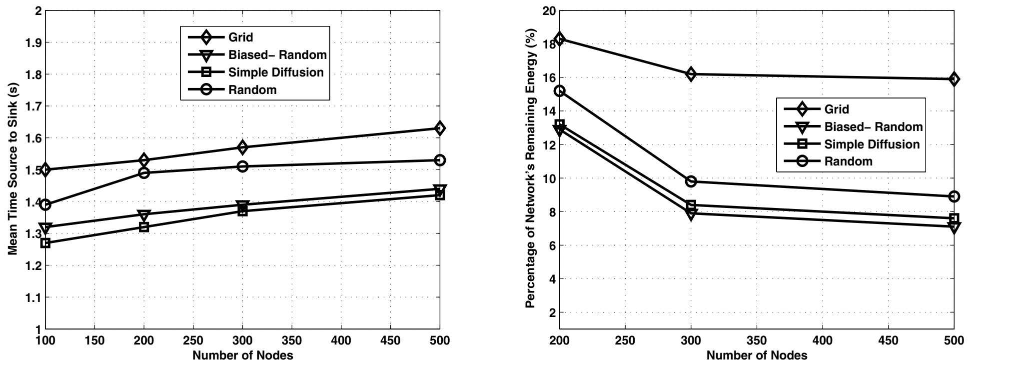
Concerning the average data rate, Fig. 9a shows that it is affected by the topology. The Biased-Random placement scheme, exhibits the best performance followed by Simple Diffusion and Random Placement. The worst performance is exhibited by Grid Placement. The reason is that in Biased Random Placement, in a higher grade and Simple Diffusion in a lower, there is a high concentration of nodes one hop away from the sink. This leads to the creation of multiple disjoint "good" paths, which can be reinforced by the algorithm, to constantly maintain high data rates. On the other hand, in random placement as well as in grid placement, the nodes around the sink are limited. This leads to few high performance paths and the data rate is kept low.

Packet drops lead to higher latency paths which are not desirable especially in WSNs. Directed Diffusion uses negative reinforcement to prune off higher latency paths.

In Fig.9b we record packet drops in all four topologies. It is clear that Simple Diffusion placement, due to the plethora of paths that constantly reinforces, presents null packet drops (after the initial injection of data packets in the network). On the other hand on the other topologies there are some packet drops, but negative reinforcement handles them quickly and efficiently.

Due to the nature of Directed Diffusion there is not much deviation between the four topologies, concerning mean packet's delivery delay. The reason is the reinforcement of high performance paths and negative reinforcement of low performance paths, which allows to the network to prune off high latency paths. In spite of this, Simple Diffusion followed by Biased-Random placement exhibits the best performance, in comparison with the other placements (Fig.10a), since it achieves nearly null number of packet drops (after the first injection of data packets in the network).

Directed diffusion presents much better performance compared to ESRT and SenTCP concerning the percentage of network's remaining energy. Comparing the performance of Directed Diffusion in different placements (Fig.10b) we record again, that when algorithm runs on placements like Simple Diffusion and Biased- Random, which are capable to provide many paths from source to sink, manages to utilize the network's resources uniformly.



(a) Directed Diffusion:Mean Packet's Delivery Delay (b) Directed Diffusion:Percentage of Network's Remaining Energy

Figure 10. Directed Diffusion:Mean Packet's Delivery Delay and Percentage of Network's Remaining Energy

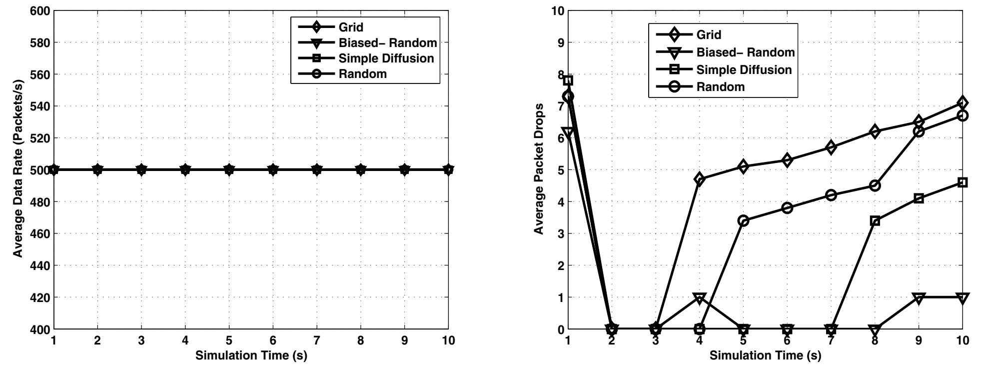
Good performance is also presented with random placement, since Directed Diffusion finds multiple paths for forwarding the data from source to sink. On the other hand, placement like Grid, are more vulnerable to the creation of network holes, due to the limited number of nodes near sources and sinks.

6.5. HTAP evaluation

By its nature, HTAP algorithm does not control the data rate of the sources, since it is an algorithm that mitigates congestion through the employment of sleep nodes ("resource control"). As it is expected, the network's data rate is kept stable in all simulation moments and for all topology schemes (Fig.11a).

Fig.11b presents HTAP's average packets drops. We observe that Biased- Random Placement exhibits the fewer packets drops compared to the other placements. The reason lies on the big number of nodes around sources and sinks. In this algorithm, in which the data rate is not reduced, the existence of many nodes one hop away from source and one hop away from sink is very important. Otherwise, if the nodes around sources and sinks are limited, the network will experience "hot-spot" congestion, at these nodes. This is what is happening in Random and Grid topology. Grid and Random placements face this situation very soon, since the number of nodes around the sources and sinks is limited, while Simple Diffusion faces this situation later due to the larger number of nodes (compared to Random and Grid topology) around the sink.

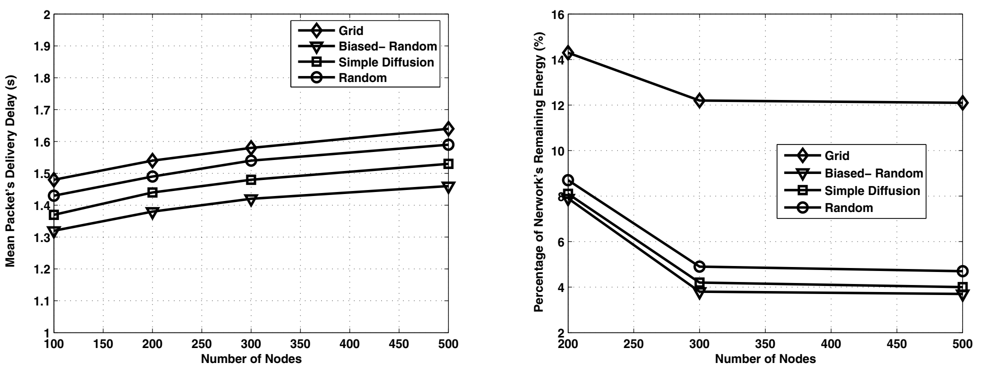
In Fig.12a we observe that there is a deviation in mean packet's delivery delay between the four topologies, as the node density increases. Biased- Random and Simple Diffusion placements, seem to cope better with the increment of the number of nodes, as this increment creates more data paths. The reason in this case, as well, is the number of nodes around the sink, nodes that can directly forward the data to sink. On the other hand, as fewer nodes are around sink, latency increases due to the "hot spot" that appears near sink.



(a) HTAP: Average Data Rate (packets/s)

(b) HTAP:Average Packet Drops

Figure 11. HTAP:Average Data Rate (packets/s) and Average Packet Drops



(a) HTAP:Mean Packet's Delivery Delay

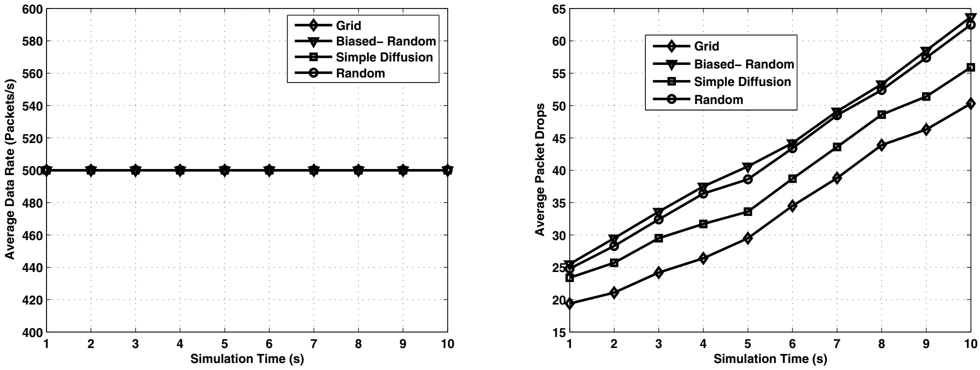
(b) HTAP:Percentage of Network's Remaining Energy

Figure 12. HTAP:Mean Packet's Delivery Delay and Percentage of Network's Remaining Energy

Concerning the percentage of network's remaining energy (Fig. 12b), we notice that HTAP exhibits the best performance compared to the other algorithms. The reason is the creation of alternative paths, that employs almost all nodes in the procedure of packets forwarding from source to sink. Comparing the performance of HTAP in the four placements we notice that beside Grid Placement, on the other three placements, HTAP algorithm exhibits very good performance and utilizes more than 90% of network resources. This is a strong indication that "resource control" algorithms can significantly increase the lifetime of a heavy loaded network.

6.6. Flooding evaluation

Flooding is generic routing algorithm. When flooding applies, each node forwards every message to every node that is in its radio range. Since it does not implement any "traffic control" functionality in case of congestion, then the sources data rate remains the same (Fig. 13a).



(a) Flooding: Average Data Rate (packets/s)

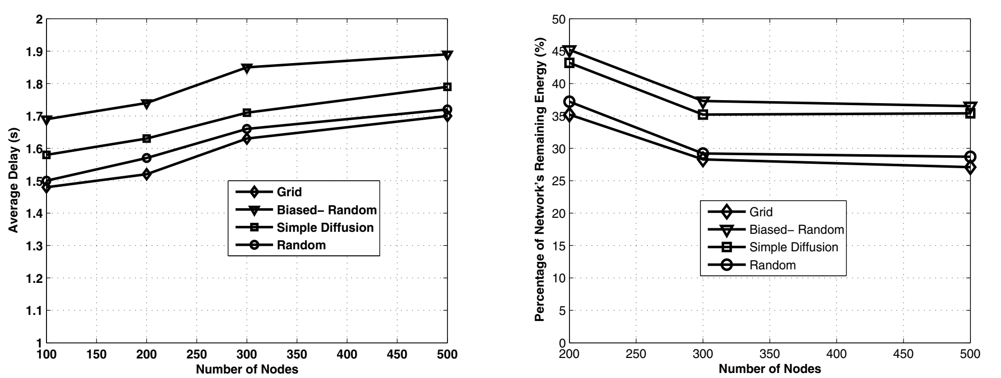
(b) Flooding:Average Packet Drops

Figure 13. Flooding:Average Data Rate (packets/s) and Average Packet Drops

Fig. 13b presents average packet drops when flooding algorithm applies. As it is expected packets drops increase while time evolves. The reason lies on the functionality of flooding algorithm. That is the fact, that fills network with multiple copies of the same packet. Comparing just the placements, we notice that the placements that present the worst performance in the previous cases, when flooding applies present the best. Grid and Random are the placements with the fewer nodes around source. Counting that flooding algorithm, forwards every message to every node in its radio range, it is normal that the placements with the fewer paths, limit the number of packets in the network and thus the fewer drops appear.

The same attitude is depicted with mean packet's delivery delay (Fig. 14a). Again placements that create fewer paths are able to forward the data sooner.

Finally we check the percentage of network's remaining energy when the network stalls (Fig. 14b). In this case we also notice that placements with fewer nodes around source (Grid and Random placements) present better performance compared to the other placements. The



(a) Flooding:Mean Packet's Delivery Delay

(b) Flooding:Percentage of Network's Remaining Energy

Figure 14. Flooding:Mean Packet's Delivery Delay and Percentage of Network's Remaining Energy

reason lies on the operation of Flooding algorithm, (each node transmits each packets to all of its children). This leads to a bigger amount of transmissions from the nodes that have many children around source (this happen in Biased Random and Simple Diffusion placement) and soon these nodes are getting power exhausted. This leads to the creation of a "hole" around source and network "stalls".

7. Conclusions

In this paper we evaluated the performance of specific routing and congestion control algorithms when nodes are deployed under different placements. The algorithms we examined are ESRT ([7]), Sen- TCP ([8]), Directed Diffusion ([9]), HTAP ([10]) and Flooding, in Simple Diffusion, Random, Grid and Biased- Random Placements. Each algorithm represents a special category of congestion control and routing algorithm. ESRT is "reliable data transport" algorithm, SenTCP is a congestion control algorithm that mitigates congestion using "traffic control" method, Directed Diffusion discovers and maintains multiple high performance paths for transmitting packets from source to sink while HTAP is a congestion control algorithm that employs "resource control" method. Finally, flooding is a generic routing algorithm, that its functionality lies on the fact that each node tries to forward every message to every one of its neighbor nodes.

Simulation results show, that algorithms that employ multiple and alternative paths, either by default (Directed Diffusion) or as a response to a congestion situation (HTAP), for the transmission of data from source to sink are significantly favored by denser placements of nodes around source and sink since they can create many paths. This leads to fewer packet drops, while they extend significantly network's lifetime. On the other hand algorithms that always employ the shortest path for the transmission of packets from source are not affected by different node placements and in case of continuous heavy load they present shortest network's lifetime.

8. Future work

Node placement is proven to be an effective way, for optimizing the performance in WSNs concerning "resource", congestion control algorithms. A future work on this subject would be the application of these placements on a real WSN environment and study the performance on a real network. Moreover, it would be worth studying what placements can assist "traffic control" algorithms to increase their performance. Initial results show that placements that create short paths from sources to sinks can assist in this direction. Finally, other parameters like fault tolerance, fault recovery, etc., are possible to be affected by different node placements. Examination of these issues constitutes part of our future work on the subject.

Acknowledgment

This work has been partly conducted under the European Union Project GINSENG funded under the FP7 Program (FP7/2007-2013) grant agreement no 224282.

Author details

Charalambos Sergiou and Vasos Vassiliou
University of Cyprus,1 University Ave., 2019, Nicosia, Cyprus

9. References

[1] Alberto Cerpa, Jeremy Elson, Michael Hamilton, Jerry Zhao, Deborah Estrin, and Lewis Girod. Habitat monitoring: Application driver for wireless communications technology. In *SIGCOMM LA '01: Workshop on Data communication in Latin America and the Caribbean*, pages 20–41, New York, NY, USA, 2001. ACM.

[2] E. Biagioni and K. Bridges. The applications of remote sensor technology to assist the recovery of rare and endangered species. *International Journal of High Performance Computing Applications, Special Issue on Distributed Sensor Networks*, April 2003.

[3] Loren Schwiebert, Sandeep K. S. Gupta, and Jennifer Weinmann. Research challenges in wireless networks of biomedical sensors. In *MobiCom '01: Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 151–165, New York, NY, USA, 2001. ACM Press.

[4] H. T. Kung and D. Vlah. Efficient location tracking using sensor networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1954–1962, March 2003.

[5] R.R. Brooks, P. Ramanathan, and A. M. Sayeed. Distributed target classification and tracking in sensor networks. *Proceedings of the IEEE*, 91(8):1163–1171, 2003.

[6] B.C. Srean, J.S. Silva, L. Wolf, R. Eiras, T. Voigt, U. Roedig, V. Vassiliou, and G. Hackenbroich. Performance control in wireless sensor networks: the ginseng project - [global communications news letter]. *Communications Magazine, IEEE*, 47(8):1 –4, august 2009.

[7] Y. Sankarasubramaniam, O. Akan, and I. Akyildiz. ESRT: Event-to-Sink Reliable Transport in Wireless Sensor Networks. In *MobiHoc '03: Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing*, pages 177–188, New York, NY, USA, 2003. ACM.

[8] C. Wang, K. Sohraby, and B. Li. SenTCP: A Hop-by-Hop Congestion Control Protocol for Wireless Sensor Networks. *IEEE INFOCOM (Poster Paper)*, March 2005.

[9] Chalermek Intanagonwiwat, Ramesh Govindan, Deborah Estrin, John Heidemann, and Fabio Silva. Directed Diffusion for Wireless Sensor Networking. *IEEE/ACM Transactions on Networking*, 11(1):2–16, 2003.

[10] Charalambos Sergiou, Vasos Vassiliou, and Andreas Pitsillides. Reliable Data Transmission in Event-Based Sensor Networks During Overload Situation. In *WICON '07: Proceedings of the 3rd International Conference on Wireless Internet*, pages 1–8, Austin, Texas, October 2007.

[11] V. Vassiliou and C. Sergiou. Performance Study of Node Placement for Congestion Control in Wireless Sensor Networks. In *New Technologies, Mobility and Security (NTMS), 2009 3rd International Conference on*, pages 1 –8, dec. 2009.

[12] Charalambos Sergiou and Vasos Vassiliou. Energy utilization of HTAP under specific node placements in Wireless Sensor Networks. In *Wireless Conference (EW), 2010 European*, pages 482 –487, 12-15 2010.

[13] Mohamed Younis and Kemal Akkaya. Strategies and Techniques for Node Placement in Wireless Sensor Networks: A Survey. *Ad Hoc Networks*, 6(4):621–655, 2008.

[14] Stavros Toumpis and Leandros Tassiulas. Optimal Deployment of Large Wireless Sensor Networks. *IEEE Transactions on Information Theory*, 52(7):2935–2953, 2006.

[15] Mika Ishizuka and Masaki Aida. Performance Study of Node Placement in Sensor Networks. *ICDCSW '04: Proceedings of the 24th International Conference on Distributed Computing Systems Workshopsd*, pages 598–603, March 2004.

[16] H. Zhang and J. C. Hou. Is deterministic deployment worse than random deployment for wireless sensor networks? In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–13, April 2006.

[17] A. Woo and D. E. Culler. A Transmission Control Scheme for Media Access in Sensor Networks. In *Proceedings of the 7th annual International Conference on Mobile Computing and Networking (MobiCom'01)*, pages 221–235, New York, NY, USA, 2001. ACM Press.

[18] Sameer Tilak, Nael B. Abu-Ghazaleh, and Wendi Heinzelman. Infrastructure tradeoffs for sensor networks. In *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 49–58, New York, NY, USA, 2002. ACM.

[19] Chenyang Lu, Brian M. Blum, Tarek F. Abdelzaher, John A. Stankovic, and Tian He. RAP: A Real-Time Communication Architecture for Large-Scale Wireless Sensor Networks. Technical report, Charlottesville, VA, USA, 2002.

[20] Cheng Tien Ee and Ruzena Bajcsy. Congestion Control and Fairness for Many-to-One Routing in Sensor Networks. In *SenSys '04: Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 148–161, New York, NY, USA, 2004. ACM.

[21] Bret Hull, Kyle Jamieson, and Hari Balakrishnan. Mitigating Congestion in Wireless Sensor Networks. In *SenSys '04: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, pages 134–147, New York, NY, USA, 2004. ACM.

[22] Chieh-Yih Wan, Shane B. Eisenman, and Andrew T. Campbell. CODA: Congestion Detection and Avoidance in Sensor Networks. In *SenSys '03: Proceedings of the 1st international Conference on Embedded Networked Sensor Systems*, pages 266–279, New York, NY, USA, 2003. ACM Press.

[23] Karthikeyan Sundaresan, Vaidyanathan Anantharaman, Hung-Yun Hsieh, and Raghupathy Sivakumar. ATP: A Reliable Transport Protocol for Ad Hoc Networks. *IEEE Transactions on Mobile Computing*, 4:588–603, 2005.

[24] Fred Stann and John Heidemann. RMST: Reliable Data Transport in Sensor Networks. In *Proceedings of the First International Workshop on Sensor Net Protocols and Applications*, pages 102–112, Anchorage, Alaska, USA, April 2003. IEEE.

[25] Chieh-Yih Wan, Andrew T. Campbell, and Lakshman Krishnamurthy. P.S.F.Q: A Reliable Transport Protocol for Wireless Sensor Networks. In *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless Sensor Networks and Applications*, pages 1–11, New York, NY, USA, 2002. ACM Press.

[26] Jaewon Kang, Yanyong Zhang, and Badri Nath. TARA: Topology-Aware Resource Adaptation to Alleviate Congestion in Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, 18(7):919–931, 2007.

[27] C. Sergiou and V. Vassiliou. DAlPaS: A Performance Aware Congestion Control Algorithm in Wireless Sensor Networks. In *Telecommunications (ICT), 2011 18th International Conference on*, pages 167 –173, may 2011.

[28] The Network Simulator ns-2 (v2.1b8a). http://www.isi.edu/nsnam/ns/, October 2001.

Chapter 2

Power Optimization for Wireless Sensor Networks

A.R. Pinto, L. Bolzani Poehls, C. Montez and F. Vargas

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50603

1. Introduction

Wireless Sensor Network (WSN) is a denomination that covers a lot of variations in compositions and deployment. A typical sensor network consists of a large number of low cost, low power distributed devices, called nodes, deployed in the environment being sensed and controlled (Stankovic et al., 2003). In other words, this kind of network is composed of a huge number of tiny nodes able to communicate with each other that can be used to monitor hazardous and inaccessible areas. Thus, each node consists of processor, memory, wireless antenna, battery and the sensor itself. Nodes can sense scalars from the environment such as temperature, acoustic and light, but may also process and transmit them by radio. The network can be classified as homogeneous or heterogeneous, which would mean that some specific nodes present special hardware or software configuration, but even in homogeneous networks, to collect, store and process data from the WSN's nodes, a special node, called Base Station (BS), is necessary.

Most of the currently adopted technologies for WSNs are based on low-cost processors, resulting in limited energy budget and restricted memory space. In many applications, it is expected that the sensor node last for a long time because in most of the cases these networks are used in remote areas and recharging and/or replacing power supply units is considered difficult or prohibitive due to hazardous and inaccessible places where they are supposed to operate. Further, due to the availability of cheap hardware and various possibilities for the radio communication frequency, numerous topologies for WSN can be adopted (Akyildiz, 2002; Ilyas & Mahgoub, 2005; Oliver & Fohler, 2010).

As previously mentioned, the nodes in these networks are usually inexpensive and therefore WSNs may be composed of a huge number of sensor nodes, which themselves are deployed inside and/or around the phenomenon that one desires to monitor. Not necessarily the sensor node's geographic position is previously known in all adoptions, because, when the nodes need to work in hazardous or inaccessible areas, it might be impossible to avoid their random deployment. Thus, if this type of application is adopted, the use protocols and approaches that can self-organize and self-optimize energy consumption of a large number of nodes that cooperate in order to achieve a global goal, becomes necessary (Akyildiz, 2002). Summarizing, WSN are desired to present the following characteristics (Ilyas & Mahgoub, 2005):

- Self-organization;
- Short range wireless communication and multi-hop routing;
- Large number of nodes and cooperative efforts between WSN nodes;
- Different WSN topologies, which frequently change due to battery depletion and node faults;
- Constrained resources such as energy budget, processing and memory.

The characteristics above and the capacity of interaction with the environment distinguish WSNs from other ad-hoc wireless networks. WSNs, due to the scarce software and hardware resources, are application oriented; thus, WSN applications are developed focusing on a specific problem's solution. The collaborative efforts between nodes are necessary to the correctly use the WSN's resources, this effort can also increase the WSN's lifetime (Stankovic et al., 2003; Hadim & Mohamed, 2006).

It is important to highlight that the strategy of deploying a large number of unreliable nodes presents advantages when compared with deploying few expensive but very reliable nodes (Ilyas & Mahgoub, 2005). These advantages are listed below:

- Larger spatial resolution;
- Higher fault tolerant degree achieved throughout distributed techniques;
- More uniform coverage;
- Ease of deployment;
- Reduced energy consumption;
- Increased network lifetime.

WSNs present a high degree of environmental interaction, depending on where sensor nodes are deployed, implicit and explicit temporal restrictions apply. In this context, data freshness is an important concept that dictates how long a sensed scalar can be considered useful and when it can be discarded. In the following example, the information gathered by a security application based on WSN technology, identifies any person who enters into a certain area of the building in a certain time, all data that exceeds this time limit, is not useful. Despite having time constraints, due to the high node density, non-determinism, noise and constrained WSN resources, it becomes extremely difficult to guarantee real-time properties (Stankovic et al., 2003). Therefore, these special constraints impose that no hard deadlines are considered for WSN application. Consequently, critical real-time systems are out of scope for this kind of network (Koubaa et al., 2009; Oliver & Fohler, 2010).

WSN can be considered an innovative paradigm, which permits the emergence of several new monitoring applications, but introduces challenges intrinsic to this technology. Some of these challenges are (Akyildiz, 2002; Stankovic et al., 2003; Molla & Ahamed, 2006; Yick et al., 2008; Ilyas & Mahgoub, 2005):

- *Paradigm change:* WSN are basically deployed in order to collect scalars from the environment and support control applications. The WSN application must sense the environment and, sometimes, act in one way or the other on the environment. Thus, it is considered critical to obtain a cooperative behaviour of thousands of sensor nodes where the data from just one node may not be important. The sensor nodes do not have a permanent identification address, due to the fact that generally messages are not sent to a specific node but to a space or area (based on the message's content). Users can be interested in the information about a specific monitoring area, thus the sensed data from a specific node may not be important, representing the data centric approach in WSNs. The need of physical environment interaction also implies important differences between WSNs and ordinary ad-hoc networks, thus classical distributed system techniques are not applicable to WSNs. Real-time requirements, noise, high fault occurrence and non determinism also impose a new group of approaches that must deal with these constraints (Molla & Ahamed, 2006).
- *Resource Constraints:* As noted above, WSNs face severe resource constraints. The main resource constraints are: limited energy budget, restricted CPU clock, restricted memory as well as network bandwidth. These characteristics impose the application of new solutions. The fact that WSN topologies are composed of a huge number of nodes represents a new issue that had not been considered in simple ad-hoc networks. For instance, trade-off approaches that aim at guaranteeing energetic economy and real-time characteristics became necessary (Yick et al., 2008).
- *Unpredictability:* There are many uncertainties that may affect a WSN. Firstly, WSN are deployed in environments with multiple uncontrollable events. Secondly, wireless communication is sensitive to noise; data lost due to radio interference and several physical errors is common to networks employed in harsh environments. Thirdly, nodes are not individually dependable. Further, it is not always possible to properly calibrate the nodes before employment; routing structures such as paths and the connectivity can be dynamically added or excluded during the WSN time of functioning. The addition or removal of nodes might be necessary due to permanent faults or battery depletion. Additionally, the energy level in some node can significantly vary even during the initial deployment. Last, the nodes might be physically removed due to environment causes or intentional controlling, thus a network restructuring must would be necessary.
- *Self-*:* One of the biggest challenges is to create the WSN's vision in the network application layer. Due to the fact that WSNs are deployed to operate with few or none human intervention, self-* characteristics like self-organization, self-optimization and self-healing become necessary (Huebscher & McCann, 2008; Oliver & Fohler, 2010). These characteristics are easily listed as challenge, however are extremely difficult to achieve.

- *High scale/density:* There are several WSN approaches that consider a large number of nodes in order to overcome hardware or software faults, thus there is a minimum number of nodes that is necessary to guarantee the WSN's service. The main challenges include: the processing of this large number of generated data, the assurance that the particular WSN requires the minimum desirable density, and the development of solutions that require the lowest density and energy consumption in order to maximize the WSN's lifetime. A WSN based on a large number of nodes that are deployed in large areas is considered a large-scale system. Due to its characteristics, these systems are subject to faults, noise, which sometimes can be caused by the WSN itself, and other uncertainties. Moreover, when a WSN is deployed, it might be self-operational and present self-maintenance, due to the fact that human intervention is sometimes very expensive or even impossible. Therefore, all these characteristics impose several conflicting goals. These challenges can be increased by the technology scaling, where the industry's minimization tendency (Akyildiz et al., 2008).
- *Real-time:* WSNs operate in the real world, thus real-time features are necessary to guarantee the correct functionality. These systems present implicit real-time constraints. The response time of its tasks is also important, thus the system tasks must be finished as fast as possible. Several WSNs present explicit real-time constraints. For example, a structural monitoring application imposes explicit deadlines for the data sensing (Kim et al., 2007). However, due to the large number of nodes, non-determinism and noise, it might be extremely hard to guarantee real-time properties.
- *Security:* WSNs can be used in safety critical applications, thus their security is an essential issue to be considered. Denial of Service techniques can be easily executed over a WSN. Moreover, coordination and real time communication approaches do not consider security issues. Thus, some intruder can easily exploit these WSN security faults. The great dilemma is how to implement security techniques that need large computational resources in a technology that deals with severe hardware constraints.

In this scenario, where nodes are likely to operate on limited resources, power conservation is considered one of the most important concerns of these networks and different strategies and protocols need are adopted in order to deal with it (Gholamzadeh & Nabovati, 2008). In more detail, network lifetime can be enhanced if the system's software, including different layers and protocols, is designed in a way that lowers the consumption of energy (Gholamzadeh & Nabovati, 2008). Several techniques have been proposed in literature in order to decrease the power consumption of WSNs. These techniques are related to different aspects of sensor networks, from hardware platform to Medium Access Control (MAC) protocol, routing and topology control.

This Chapter is structured as follows: Section 2 summarizes the main applications where WSNs are deployed and their hardware characteristics. In Section 3, the main MAC layer approaches proposed in literature are described. Section 4 presents the routing strategies proposed in order to provide power optimization and consequently increasing WSN's lifetime, while Section 5 introduces Transmission Power Control approaches. Section 6 introduces Autonomic approaches and finally, in Section 6 the final remarks on the optimization of WSNs are presented.

2. WSN applications and hardware characteristics

WSNs are considered an application oriented technology, thus approaches that are developed for some specific application usually cannot be used for different uses. Important points related to the hardware characteristics of the nodes must be considered in order to guarantee the suitable node for a specific application. In more detail, aspects related to the type of processing unit as well as communication, power supply and sensing devices must be considered, when the nodes for a specific application are defined.

Considering aspects related to the processing unit, usually a microcontroller or microprocessor is adopted. In order to choose the ideal microcontroller for the system and due to the fact that microcontrollers with high performance imply higher power consumption, the designer must consider the desired performance level. Another important point is associated to the fact that microcontrollers usually support different operational modes, such as active, idle and sleep mode, which directly affect the power consumption of the node. There is also an attractive design option that suggests splitting the workload between two low power microcontrollers in such way that one of the microcontrollers is responsible for the sensing control, while the other performs the networking tasks related to controlling the RF interface and running the algorithms (Chou & Park, 2005). Finally, techniques like the one called Dynamic Volta Scaling (DVS) can be adopted (Karl & Willig, 2005). DVS dynamically adapts the microcontroller's power supply voltage and operating frequency to meet the processing requirement, thus trading off performance and power supply for energy savings.

Different communication devices using mediums like radio frequency or optical communications, for example, can be adopted to exchange data between nodes. For communication, both a transceiver and a receiver are required for the sensor nodes. The essential task of these devices is to convert a bit stream coming from a microcontroller into radio waves and vice versa. In more detail, the transceiver is normally regarded the largest power consumer and optimizing its power can result in significant improvement for the system as a whole (Chou & Park, 2005). There are several factors that affect the power consumption characteristics of a transceiver, including its type of modulation scheme, data rate, transmission power and the operational duty cycle (Gholamzadeh & Nabovati, 2008). Many transceivers allow the user to set the power level. In general, transceivers can operate in the following distinct modes of operation: Transmit, Receive, Idle and Sleep; allowing switching between them and consequently realizing energy savings. Note that the switching between the operating modes has to be managed, taking into account the fact that waking up a transceiver from the Sleep mode and making it go to Transmit mode requires some start-up time and start-up energy. Thus, switching a node into Sleep mode is only beneficial, if the energy necessary for the node to comeback into an active mode is smaller than the energy saved during the Sleep mode, which implies that the time to the next event is sufficiently large.

Regarding the power supply device, a battery is used in most of the cases, playing a vital role in determining the sensor node's lifetime. Thus, one of the most important factors that a designer must consider is the rate capacity effect, which is related to the discharge rate or the amount of current draw from the battery. Drawing higher current than the rated value leads to a significant reduction in battery life, due to the fact that the diffusion of electrolyte falls behind the rate at which they are consumed at the electrodes. It is important to highlight that most of the applications of WSNs involve deploying sensor nodes in harsh and remote environment and therefore it is difficult to use ordinary recharging schemes for batteries. In particular cases, an alternative is to adopt external energy resources like sunlight or wind.

Finally, sensors in WSNs translate physical phenomena to electrical signals and can be classified as analog or digital devices depending on the type of output they produce. Basically, there are several sources of power consumption in a sensor: signal sampling and conversion of physical signals to electrical ones, signal conditioning, and analog to digital conversion (Gholamzadeh & Nabovati, 2008). Passive sensors, such as temperature sensors, consume less power than active sensors, like sonar, which need energy to send out a signal to probe the observed object. Indeed, the sampling rate is important and higher frequency sampling requires more energy. In this context, sensors should acquire a measurement sample only if needed, when needed, where needed and with the right level of fidelity (Raghunathan et al., 2006). This strategy reduces the energy consumed in the subsystem and sometimes reduces the processing and communication load as well. Thus, the use of mechanisms able to change the bit resolution of measurement samples and the sampling rate as well as using adaptive spatiotemporal sampling, exploiting redundancy and correlations models to predict a measurement instead of actually making it and finally, hierarchical sensing, can provide power consumption optimization.

In the next paragraphs, several WSN applications will be briefly presented. According to the temporal requirements of the applications, they may present significant differences in the applied algorithmic solutions. For instance, an environment monitoring application can require deadlines of minutes, while in a military application temporal validity is much smaller. Some kinds of applications need periodic sensing and sending, while other applications need an event driven approach.

Possible applications of WSNs include environment monitoring, military, domotic and industrial monitoring and control. For instance, an application of habitat monitoring is presented in (Mainwaring et al., 2002). The presented WSN has been deployed on the Great Duck Island. Its main goal is to correlate the measurements of some microclimate data (temperature, light and humidity) with the bird nest activity on the island. This application presents relaxed real-time requirements, thus the main goal of the Great Duck Island application is the maximization of the network's lifetime, as it is expected that the WSN's infrastructure stays active during months or even years without human intervention. Therefore, the intervals between sending messages and between one sensing and another may be reduced significantly.

Another example for a WSN's application is structural monitoring, in this case a linear WSN topology was used to monitor the Golden Gate Bridge's structure, and thus a routing technique had to be applied to assure the messages delivery to the BS at one end of the construction. This application is based on accelerometer sensors that detect modifications in the physical structure of the bridge (Kim, S. et al., 2007).

Finally, other kinds of WSNs' utilization are stated below:

- *Automotive industry:* Cars are equipped with sensor and actor networks, which can interact with highway or street WSN infrastructures in order to increase traffic efficiency or automate toll payment;
- *Monitoring and automation of factory systems:* Industrial robots can be equipped with thousands of wired sensors. These sensors must be connected to a central computer. The high economic cost and the mobility restrictions of wired sensor are favour the utilization of WSNs in this kind of robots.
- *Intelligent housing:* WSNs permit that houses can be equipped with movement, light and temperature sensors, microphones used for voice activation and pressure sensors in chairs are also examples of WSN utilization in building automation. Thus, air temperature, natural and artificial lighting and other components can be tuned according to specific user needs;
- *Merchandise tracking:* Logistic and transportation companies may use WSN technology in order to track ships, transporters, containers or single goods that are being transported;
- *Precision agriculture:* Irrigation control and precise pesticide application are possible with the help of WSN utilization on farmlands;
- *Harsh area monitoring:* Exploration and monitoring of harsh areas may be possible throughout the use of WSNs;
- *Freshwater quality monitoring:* WSNs may be used for freshwater monitoring due to their non-intrusiveness and small size;
- *Military application:* Position and movement control of troops and vehicles, target detection, non-human combat-area monitoring as well as landmine removal or building exploration are just some examples of possible utilizations of WSNs for military applications.

To conclude, WSNs can be applied in different types of applications and the selection of the suitable hardware depends on the systems' requirements, the available resources and the environment where the network should operate.

3. MAC layer approaches

As previously mentioned, the lifetime maximization of WSNs is one of the most important concerns when dealing with the use of WSNs. This is mainly related to the fact that sensor nodes are considered unavailable when the battery level is depleted. In this context, it is important to note that communication among nodes is the major energy consumer process in WSNs. A significant portion of the node's energy is spent on radio transmissions and on listening to the medium for anticipated packet reception (Gholamzadeh & Nabovati, 2008). In other words, sending or receiving messages requires significantly more energy than data processing or the acquisition by the sensors. Moreover, a single medium for communication is shared by the nodes and network performance largely depends on how efficiently and fairly these nodes share the medium. MAC protocol controls the communication nodes in WSNs and regulates access to the shared wireless medium such that the performance requirements of the underlying applications are satisfied (Sohraby et al., 2007). Thus, a careful definition of protocols and algorithms for efficient communication has become one of the most important issues in WSNs in order to improve their lifetime. Basically, the MAC protocol must be energy efficient and must try to reduce the following issues related to energy consumption phenomena (Demirkol, 2006):

- *Packet collision*: When one node receives more than one packet at the same instant, it is considered that a packet collision occurred. Therefore, all packets must be discarded and transmitted again.
- *Overhearing*: When some node receives packets that are addressed to another sensor node overhearing occurred;
- *Control packet overhead*: The use of control packets in order to coordinate the WSN must be minimized;
- *Idle listening*: Idle listening occurs when some node is in the listening mode of a channel that is not being used.
- *Over-emitting*: Over-emitting is caused when the message delivery fails due to the destination node's inactivity.

It is important to highlight that the collisions of messages is considered the most critical aspect, which causes the discarding of all involved messages and forces the network to retransmit increasing its energy consumption. Thus, an energy-efficient MAC protocol must avoid collisions and reduces the energy dissipation related to idle channel sensing, overhearing and overhead to a minimum (Ilyas & Mahgoub, 2005).

Regarding the types of communication patterns, four different types can be identified for WSNs (Demirkol, 2006):

- *Broadcast*: Generally BSs use the broadcast communication pattern (sink) to transmit certain information to all nodes under its controls. The broadcast information must contain consults, software upgrades or some control packet. The broadcast pattern can only be used, when all destination nodes are inside the radio coverage of the transmitter node;
- *Local gossip*: Local gossip is to be considered when nodes sense some event and send it to other nodes that are located inside the same location (same cluster). This kind of communication occurs when one node sends its messages to its neighbours, inside the same coverage area;
- *Convergecast*: This type of pattern is used when a group of sensors sends their packets to a specific node. The destination node can be a cluster-head, a fusion centre or a BS;
- *Multicast*: Some scenarios imply that messages need to be sent to a group of sensor nodes, thus just the sensors that belong to this group receive the message.

In the next subsection the IEEE 802.15.4 Standard and the ZigBee technology are described and the main MAC approaches able to reduce power consumption are summarized as well. Another subsection, will present further MAC approaches aiming at the optimization of MSNs.

3.1. IEEE 802.15.4 standard and the ZigBee technology

The main goal of the ZigBee technology is to enable WSNs composed of large number of nodes to function with reduced energy consumption. Most WSN technologies like *Mica Motes* use ZigBee in order to achieve higher lifetime levels for their WSN applications.

The ZigBee network architecture is based on the *Open Systems Interconnection* (OSI), however exclusively the more important layers were implemented. ZigBee adopts the IEEE 802.15.4 standard, which only defines the lower layers: the physical layer and the MAC layer (ZigBee, 2010).

The physical layer may operate on two frequencies 868/915MHz or 2.4GHz, with 16 channels and 250Kbps of maximum transmission rate. The IEEE 802.15.4 MAC layer is based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA) mechanism. Note that ZigBee technology differs from other wireless technologies due to several reasons: lower data transmission rate, lower energy consumption, lower cost, higher self organization and more flexible network topologies (ZigBee, 2010).

The IEEE 802.15.4 standard has been proposed in 2003 and has become a *de facto* standard for low energy consumption and low data rate transmission networks. The IEEE 802.15.4 MAC protocol supports two kinds of operational modes that can be selected by a central node called Personal Area Network (PAN) coordinator. The two modes are:

- *Beaconless mode*: where MAC protocol functions are based on a CSMA/CA without beacon packet;
- *Beacon mode*: where beacons are periodically sent by the PAN coordinator in order to synchronize nodes that are associated with it and to delimit a *superframe. D*uring the *superframe* duration all node transmission must occur. Moreover, during the contention period of this frame the MAC protocol is ruled by the slotted CSMA/CA. The IEEE 802.15.4 with beacon mode can use the synchronization and the contention free period that is based on a guaranteed slot time.

Thus, the ZigBee Alliance is responsible for the ZigBee technology standardization. In more detail, the application and network layers are defined by the ZigBee Alliance itself, while the physical and MAC layers are based on the IEEE 802.15.4 standard. ZigBee may also consider time synchronization according to an optional *superframe* structure; the ZigBee technology possesses an address scheme that can handle up to 65.000 nodes. Moreover, three kinds of topologies are supported: *star*, *mesh* and *cluster tree*. The *Star* topology is considered the simplest topology and is based on a many-to-one communication topology, which means that all nodes are covered by the PAN coordinator antenna, and are able to send the messages in just one hop.

However, the *mesh* and *cluster tree* topologies rely on a routing protocol in order to deliver the messages to the PAN coordinator. *Mesh* topology does not allow cluster-heads and nodes to communicate with each other. Still different, the *cluster tree* topology is based on the organization of nodes into clusters. Basically, the *star* topology is considered simpler than the *mesh* and *cluster-tree* ones due to the fact that no routing protocol is necessary (ZigBee, 2010).

IEEE 802.15.4 standard is based on CSMA/CA MAC algorithm, its beaconless mode does not impose the sending of a periodical beacon by the PAN coordinator (IEEE 802.15.4, 2008).

Two parameters are considered in the beaconless mode: the first is denominated NB that is the Number of times CSMA/CA is required to backoff, the second is called Backoff Exponent (BE), standing for the number of backoff periods a device must wait until it can access the communication channel.

The first step of the CSMA/CA algorithm is the initialization of NB and BE. After the initialization, the MAC layer must wait a random period of 0 to (2^{BE} - 1) and then require the Clear Channel Assessment (CCA) from the physical layer. When the channel is considered occupied by other device, the NB and BE is incremented by 1 by the MAC layer, though the MAC algorithm must guarantee that BE never grows above *macMaxBE*. Moreover, when NB´s value is above *macMaxCSMABackoffs*, the CSMA/CA algorithm must quit and returns the access channel failure status.

Finally, the Beaconless CSMA/CA algorithm is sensitive to three parameters: *macMaxBe* (standard value 5), *macMaxCSMABackoffs* (standard value 4) and *macMinBE* (standard value 3). These standard values for the parameters may help to decrease energy consumption due to the fact that devices try to send just five times before the transmission's abortion, however incrementing these values tends to increase the network's communication efficiency. Thus, the IEEE 802.15.4 protocol is not able to deal with dense network topologies, them being networks that are based on a large number of nodes.

3.2. Other MAC approaches

Several other MAC approaches have been proposed in literature in order to provide the reduction of power consumption in WSNs. In the next paragraphs the main solutions that explore the optimization of MAC protocols are summarized.

Techniques used in the MAC layer of WSNs often involve the use of Time Division Multiple Access (TDMA) and Duty Cycles (DC).

The main idea behind the TDMA technique is to divide the time spent by devices over channel accesses into so called time slots, each one used exclusively by one device. Therefore, by applying this technique, every device, before sending any messages, needs to book such a slot of time in advance. A TDMA MAC protocol is proposed in (Shi & Fapojuwo, 2010). This technique is based on a cross-layer optimization involving MAC and physical layers. The main goal of the presented technique is to reduce the overall energy consumption based on a TDMA scheduling with the shortest frame length for clustered WSNs.

In order to also reduce energy consumption, the TDMA MAC protocol is used in (Wu, Y. et al., 2010). Here, the main focus is to schedule the sensor nodes with consecutive time slots at different radio states, such as: transmitting, receiving, listening, sleeping, and idle. Due to the fact that sensor nodes consume different levels of energy at each state the optimum scheduling of these states could achieve the reduction of energy consumption.

FlexiTP is a TDMA MAC protocol that schedules node messages based on the so-called sleep scheduling approach (Lee et al., 2008). The sleeping scheduling scheme requires that sensor nodes to exclusively transmit and receive packets at their own time slot and turn into sleep state until their slot's turn is up again. FlexiTP also provides routing, time synchronization tasks and sensor nodes may sense as well as route data.

PEDAMACS is another TDMA MAC protocol designed for multihop WSNs (Ergen and Varaiya, 2006). It can improve the network's lifetime by several years when compared to other MAC protocols such as random access protocols that may reach months or just days of network lifetime. However, this TDMA protocol does not present a good performance when applied to WSNs with dynamic topologies, as they are common in harsh environments.

Complementarily, the DC technique divides the operating time of devices in two periods: active and inactive, also denominated sleeping time. The shorter the period of activity is in comparison with the period of inactivity, the longer the devices remains inactive and consequently achieves greater energy savings. As downside, a consequent reduction of the maximum transmission rate in the network is to be observed. If, on the one hand, TDMA enables devices to become more organized in order to avoid collisions; on the other hand, the DC technique is able to avoid cases where a node becomes simultaneously active with other nodes that had been inactive before, preventing a node to wait for messages that will never arrive, and finally avoiding the waste of energy.

These techniques generally allow the protocols to deal with collision, idle listening and over-emitting problems, but have overheads associated to sending and processing of control messages. Such extra-costs can be unnecessary in applications where the density of the network is small and where few devices transmit simultaneously. In this scenario, contention-based protocols such as Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA) seem to be more suitable. Note that the CSMA/CA MAC protocol present in the IEEE 802.15.4 standard (IEEE 802.15.4, 2008) is inefficient for large networks.

A Rotational Listening Strategy (RLS) for Wireless Body Network (WBN) is presented in (Tseng et al., 2011). WBNs are a special kind of WSNs that are deployed over a human body area in order to sense and transmit scalars as for example the body's temperature. The RLS approach is based on the division of channel access partitioning it into mini-slots that are allocated to nodes.

Another type of WBN application is presented in (Omeni et al., 2008), where a MAC protocol implemented in hardware by a 0.13◎m CMOS process is described. In order to avoid collisions with nearby transmitters in a wireless body network, a standard listen-before-transmit technique is being used. The time slot overlap handling was reduced based on a wakeup-fallback approach.

A different Wireless Body Networks (WBN) MAC protocol is proposed in (Otal et al., 2009). The main goal of this MAC protocol is the maximization of the battery lifetime of each individual body sensors while maintaining the reliability and message latency of data transmissions at the same level. To do so this MAC protocol is based on a cross-layer fuzzy rule scheduling algorithm and a energy-aware radio activation policy for realistic medical applications.

An Energy-Aware Hybrid Data Aggregation (EDHDAM) technique is presented in (Kim, M. et al., 2011). It aims at minimizing the energy problem in asynchronous MAC-based WSNs. The nodes closer to the sink spend more energy than other, this is due to the fact that they receive and send more data to the WSN's sink then nodes that are far away from the sink. Thus, the EDHDAM technique is designed to adaptively control the number of data transmissions in order to avoid the before mentioned downside.

A game-theoretic MAC approach for WSNs is presented in (Zhao et al., 2009). The MAC of nodes in this technique is based on an incompletely cooperative game mode. This approach denominated G-MAC, where time is divided into super-frames, each super-frame having two parts: an active part and a sleeping part. During the sleeping part, all nodes turn-off their radios to save energy and during the active part, if some node has packets to send, these will pass on the channel that is based on the incompletely cooperative game.

Multiple cross-layer protocols that integrate MAC and WSN's network layers are presented in (Rossi & Zorzi). All these MAC protocols are cost-aware regarding residual energies, link conditions, and queue state. The routing layer chooses the best relay candidates based on the MAC protocol information. In this manner, the number of in-range devices, that compete for one channel as well as the interference are reduced.

A technique named S-MAC, a medium access control based on coordinated adaptive sleep scheduling, is presented in (Ye at al, 2004). S-MAC tries to avoid the overhearing problem by low-duty-cycle operations in a multi-hop WSN. The S-MAC approach organizes the sensor nodes into virtual clusters based on common sleep scheduling in order to reduce the control overhead and enable traffic-adaptive wake-ups.

Finally, MRMAC is a MAC protocol that reduces the end-to-end delay as well as the energy consumption in WSNs. The approach presented in (Hong *et al.,* 2010) reduces the end-to-end delay based on two metrics: Next Packet Arrival Time (NPAT) and Medium Reservation Information (MRI). When a sender transmits a packet, the NPAT and MRI metrics are enclosed in the packet in order to make possible that its intended receiver reserves the medium. The simulations presented by show that the MRMAC approach is able to significantly reduce, both end-to-end delay and energy consumption.

4. Routing approaches

WSNs can be adopted for a wide range of applications, but in all of them, the main task of the nodes is to sense and collect data, process it and transmit the information to the site where it is possible to analyse the monitored parameters. To efficiently achieve this task, it is required the development of an energy efficient routing protocol to set up paths between the nodes and the data sink (Sohraby et al., 2007). Due to the fact that sensor nodes are energy constrained, great part of the WSN's protocols aim at minimizing the energy required for communications. Basically, the environment characteristics coupled with scarce resources and energy limitation make the routing problem very challenging. Thus, the path selection must be such that the lifetime of the network is maximized. In this context, different strategies can be adopted in order to face with this problem. One simple strategy is to avoid bad-quality routes because unreliability of wireless links has an adverse effect on their performance. Link failures and packet losses lead to many retransmissions and therefore, result in higher power consumption.

A clustering protocol called REACA is presented in (Quan et al., 2007). REACA's functioning is divided into two cycles: the first cycle is dedicated to the network configuration while the second cycle handles the message transmission. The cluster-head to be chosen is based on the battery level of all nodes that compose the cluster. Thus, the node that presents the highest battery energy level is chosen to function as cluster-head. Moreover, a routing algorithm is proposed and REACA is validated by mathematical analysis.

EARQ by (Heo et al., 2009) is a routing protocol based on the WSN's energy level. EARQ is able to guarantee dependability, temporal constraints and energy economy. The main goal of EARQ is to use the path with the greatest energy level inside a WSN. Its authors prove by simulations that EARQ may be implemented in a WSN for industrial application. However, EARQ was not validated for any WSN prototype. Another cluster-tree WSN routing protocol was proposed by (Alippi et al., 2009).

The approach called MMSPEED is a routing protocol that is able to guarantee probabilistic Quality of Service (QoS) metrics in WSNs. It was proposed by (Felemban et al., 2006), considers different options of speed delivery in the time domain and guarantees package delivery. Several dependability requirements are provided, which are based on several path options. The end-to-end requirements are provided in a located fashion; this is desirable in terms of scalability and adaptability in dynamic and dense WSNs. However, the utilization of geographical routing imposes that nodes need to know their geographical localization. Thus, the proposing authors considered that each node would possess GPS devices or distributed localization algorithms. This results in considerable problems, as GPS devices do not work properly in indoor environments and distributed localization algorithms impose an extra overhead due to the extra package exchange, since the nodes need to periodically broadcast their geographical localization.

The q-Switch, a simple path routing algorithm proposed in (Wu, X. et al., 2008), is a routing technique used to support the non-uniform node distribution strategy that is used to mitigate the energy hole problem in WSNs. Its authors also show that in a circular multi-hop WSN with non-uniform node distribution and constant data sending the unbalanced energy consumption is unavoidable.

An approach denominated the Energy Efficient Broadcast Problem (EEBP) in ad hoc wireless networks is presented in (Li et al., 2004). The EEBP's idea can be described by the following phrase: in a given an ad hoc wireless network, find a broadcast tree such that the energy cost of the broadcast tree is minimized. Its authors consider that all the network´s nodes present a fixed level of transmission power. As solution three routing approaches aiming at the minimization of the network´s consumption are proposed.

A sleep scheduling solution called Green Wave Sleep Scheduling (GWSS), which has been inspired by synchronized traffic lights, is presented in (Guha et al., 2011). The main goal of this approach is to support the WSN´s routing duty cycling. A green wave is a moving sequence of consecutive active states (green lights), and some packet may move in a sequence of active nodes. Thus, nodes in sleep mode are compared to red lights, and packages may not be routed through a sleeping node. Its authors show that, considering large WSNs arranged in structured topologies, GWSS achieves almost the same end-to-end latency as that of non-sleep-scheduling WSNs.

5. Transmission power control approaches

Power conservation is so important because nodes are usually operating on limited batteries. As previously mentioned, MAC protocols are able to manage energy consumption during WSN communication, which is the most energy-consuming event in WSNs. However, one interesting solution in order to increase WSN's lifetime is based on adjusting its nodes' transmission power. On the one hand, maintaining the lowest possible transmission power represents a interesting solution in order to minimize the energy consumption and consequently increase the network's lifetime. On the other hand, the lowest possible transmission power can increase the WSN's vulnerability to the interference fluctuations caused by bad Signal-to-Interference-plus-Noise-Ratio (SINR) (Kim & Know, 2008). Extensive empirical studies confirm that the radio communication's quality between low power sensor devices varies significantly with time and environment. This phenomenon indicates that the existing topology control solutions, which use static transmission power, transmission range and link quality, might not be effective in the physical world (Lin et al., 2006). In this context, online transmission power control techniques that take into account environment variations have become essential in order to address this issue.

Several Transmission Power Control (TPC) approaches have been proposed in the literature. Basically, the TPC algorithm can reduce the energy consumption and improve the channel capacity. In more detail, TPC solutions use a single transmission power for the whole network, not making full use of the configurable transmission power provided by radio hardware to reduce energy consumption or assume that each node chooses a single transmission power for all the neighbours, which is know as neighbour-level solutions. Indeed, most existing WSNs use a network-level transmission power for each node.

There are many TPC studies, which mainly focus on improving the channel capacity (Monks et al., 2001; Ho & Liew, 2007). Recently, experimental studies (Don et al., 2004)(Lin et al., 2006) have shown that TPC reduces energy consumption in low-power WSNs. In Power Control Algorithm with Backlisting (PCBL), each node sends packets at different transmission power levels to determine the optimal transmission power based on the Packet Reception Ration (PRR). In (Lin et al., 2006), an Adaptive Transmission Power Control (ATPC) algorithm is proposed in order to achieve the optimal transmission power consumption for specified link qualities. Employing a ATPC algorithm, the Received Signal Strength (RSS) and the Link Quality Indicator (LQI) for radio channels are used to estimate the optimal transmission power level, and employ a feedback-based ATPC algorithm to dynamically adjust the transmission power over time. Thus, the result of applying this algorithm is that every node knows the proper transmission power level to use for each of its neighbours and every node maintains good link qualities with its neighbours by dynamically adjusting the transmission power through on-demand feedback packets.

However, the effect produced by different inference sources must be considered when the goal is the implementation of WSNs in the physical world. Many WSN devices available on the marked operate on the 2.4GHz ISM band and are vulnerable to the interferences from other wireless networks such as the IEEE 802.11 WLANs or the IEEE 802.15.1 Bluetooth (Kim & Know, 2008). Generally, the transmission power of the WSN devices is lower than that of WLAN or Bluetooth devices. Therefore, the TPC algorithm for WSNs has to carefully consider the interferences caused by other 2.4GHz wireless devices, which can cause significant performance degradation. In this context, a practical TPC algorithm for WSNs, named Interference Aware Transmission Power Control (I-TPC) algorithm has been proposed in (Kim & Know, 2008). The I-TPC algorithm is based on the idea that each node adjusts the RSS target to provide the acceptable SINR when interferences are detected. In more detail, the I-TPC algorithm consists of two functional procedures: the two-tier transmission power control and the RSS target adjustment. Initially, the proper RSS target, which may satisfy the desired PRR is determined. Based on this RSS target, each node tries to adjust its transmission power to keep the RSS value within the upper and the lower RSS target values by using the two-tier transmission power control procedure. The net effect of this operation is that the proposed algorithm tries to reach a satisfying link quality quickly even if there are small-scale link quality variations. When the interference is detected, the RSS target and the transmission power are increased immediately by the RSS target adjustment procedure to provide an appropriate SINR.

Two different local algorithms to individually adjust the nodes' transmission power are presented in (Kubish et al., 2003). Such local approaches do not require any particular MAC protocol or dedicated protocol for route discovery. The so-called Local Mean Algorithm (LMA) implements that each node periodically sends a *life message* and all receiving nodes respond with *life acknowledge messages.* Before sending new information each node counts the received acknowledge messages and compares this number to the value set as thresholds. In the case the node received less messages then the inferior threshold, the transmission power is increased by factor A_{inc} for every node missing to achieve the threshold. If this number is in the range between the minimum and maximum threshold no changes to transmission power are made. Similarly, the Local Mean Number of Neighbours (LMN) algorithm works with life and life acknowledge messages. In addition to the LMA approach the *life acknowledge message* contains the nodes own count of neighbours. Thereby each node receives a number of messages containing a value indicating the numbers of neighbours of the sending node, then calculates a mean value from all the received messages and uses this value as well as the number of nodes that responded to its *life message* to calculate the so called *NodeResp* value to be compared with the thresholds and, if the case, the transmission power is adapted as described in the LMA technique. These two techniques are compared to fixed and global algorithms and, in the given indoor scenario, are outperforming the fixed approaches while reaching only about half the lifetime of networks employing global algorithms such as the Equal Transmission Power (ETP) algorithm. It is noted that comparing such approaches to ETP the local algorithms are almost competitive when looking at the network's confidence level and on top are scalable and easily implementable, which global algorithms are not.

Finally, a Transmission Power Self Optimization (TPSO) technique is presented in (Lavratti et al., 2012). It basically consists of an algorithm able to guarantee the connectivity as well as an equally high Quality of Service (QoS) concentrating on the WSN's Efficiency (*Ef*), while optimizing the necessary transmission for data communication in each node. The technique aims at adjusting each node to use the lowest possible transmission power while maintaining the connectivity to the WSN and the reliability of the network as a whole. This trade-off between the WSN's *Ef* and the data transmission energy consumption is evaluated in different EMI environments. Its decentralized algorithm runs on the application layer and uses an *Ef* value calculated, which adopts the number of received messages and the estimate of sent messages. This *Ef* is compared with the targeted *Ef* in order to decide about adjustments to the node's transmission power. Experimental results show that the automatic adaption presents advantages when compared with approaches using fixed transmission power. It is shown that the technique is able to guarantee the trade-off between *Ef* and power consumption. The TPSO behaviour is shown in Figure 1. It is possible to notice that the energy dissipated by a node with fixed transmission power is much higher than the energy consumed by a node running the TPSO algorithm. The session values in the x-axis represent the elapsed time.

Figure 2 shows the WSN's *Ef* and the WSN's energy consumption with respect to the WSN using the five pre-defined transmission power levels and to the WSN adopting the TPSO technique.

We can observe the effectiveness of the TPSO technique with respect to the use of pre-fixed transmission power levels. In detail, we can see this network using the maximum transmission power level reaches about 80% of *Ef*, but consumes about 50mW.s, while the WSN adopting the TPSO algorithm reaches the same *Ef* consuming only about 25 mW.s.

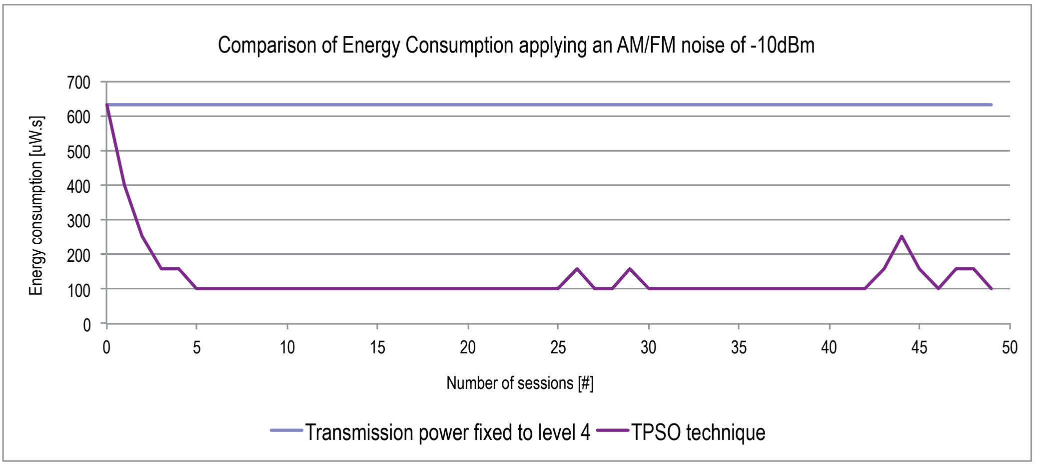


Figure 1. Comparison of the energy consumption of the TPSO technique with respect to WSN operating with the transmission power fixed to level 4 (Lavratti et al., 2012).

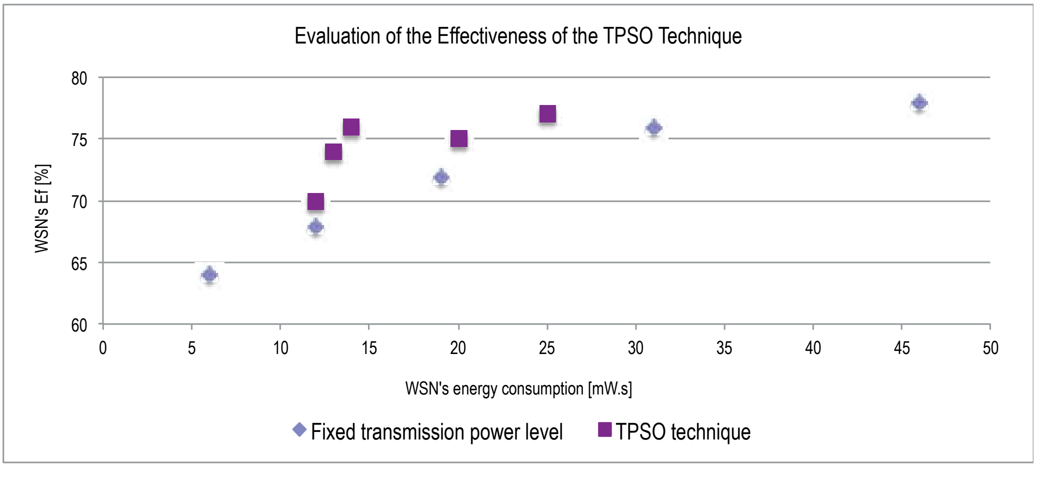


Figure 2. Evaluation of the TPSO technique with respect to WSN's Ef and energy consumption applying AM/FM noise of -14 dBm (Lavratti et al., 2012).

Figure 3 depicts the *Ef* of two WSNs, one with the transmission power level set to 0 and one with the TPSO technique. The WSN with the fixed transmission level is able to reach an average *Ef* of 46.4%, while the other network is achieving 86.6%. As the WSN with the TPSO technique is switching to higher transmission power levels to cope with the introduced noise it needs 253% more energy to reach the higher level of effectiveness.

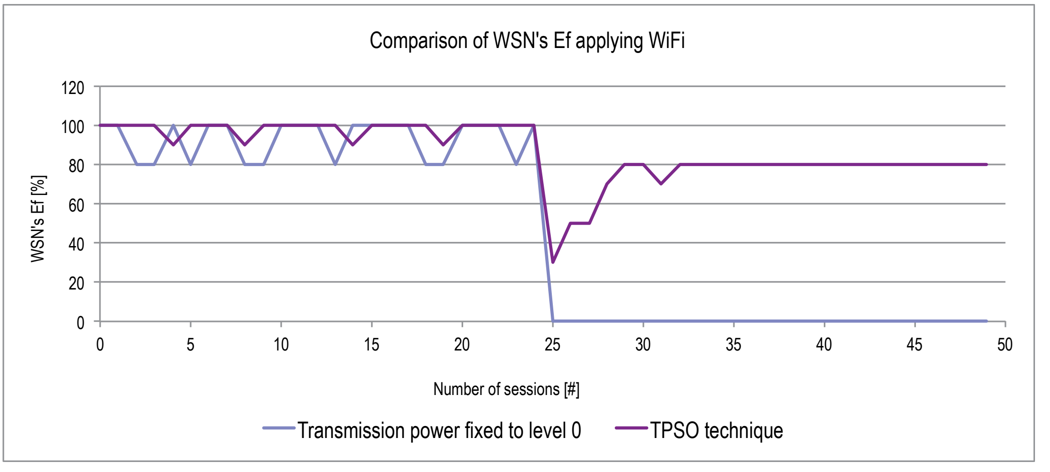


Figure 3. Comparison of the WSN's Ef of the TPSO technique with respect to WSN operating with the transmission power fixed to level 0 (Lavratti et al., 2012).

The results obtained during the experiments demonstrated the convenience of using the self-optimization algorithm instead of setting the maximum transmission power level. When a WSN without the TPSO technique is considered, the transmission power is set at the beginning of the communication and remains the same during its entire lifetime. This characteristic can be negative considering a WSN in a real environment where the inherent noise is not necessarily constant. Therefore, due to the fact that the inherent environment noise is completely variable and random, the TPSO technique will always guarantee the lowest possible transmission power during the communication and the target_*Ef* when it is possible (Lavratti et al., 2012).

6. Autonomic approaches

IBM introduced the term autonomic computing in 2001 to describe computer systems able to self-manage themselves (Kephart and Chess, 2003). The main properties of approaches categorized as "self-*" are:

- self-configuring
- self-optimizing
- self-healing
- self-protection.

Each one of them is described in (Huebscher and McCann, 2008). A brief definition is presented below:

- Self-configuration: the system's ability to configure itself according to high level goals;
- Self-optimization: the system can decide to start a change in the system as pro-active, in order to optimize the performance or quality of service;
- Self-healing: the system detects and diagnoses problems, which can be either faulty bits in a memory chip or a software error;
- Self-protection: the system is able to protect itself against malicious attacks or unauthorized changes.

Even though dense WSNs present several advantages, self-management characteristics are required in order to deal with the management of a large number of nodes. Self-management techniques are part of autonomic-computing methodologies, which can also be used to manage WSNs with conflicting targets (energy efficiency, self-organizing, time constraints and fault tolerance). The main goal of self-management is the development of a computing system that does not need the human intervention to operate. This way, computing systems are able to self-organize and self-optimize themselves, once they follow global objective dictated by a system administrator (Pinto e Montez, 2010).

For instance, in dense WSNs composed of several sensor nodes in a star network topology, in case the network presents conflicting goals (increase dependability and energy efficiency, while meeting time constraints), the conventional IEEE 802.15.4 protocol does not seem to be able to deal with the complexities. For example, when the number of nodes in a network is increased, in order to achieve better reliability the WPAN may be congested, and fewer messages arrive in the base station on time. In order to demonstrate the WSNs behaviour in this situation, experiments using TrueTime simulator[1] were performed. Two metrics called *Ef (efficiency)* and *QoF* have been adopted. While efficiency is a metric that measures the ratio between sent and received messages; *QoF* represents the average number of received messages by the base station over a certain timespan. Figure 4 shows that when density network is increased, QoF increases slowly, but communication efficiency quickly decreases.

(Pinto e Montez, 2010) propose a Genetic Machine Learning Algorithm (GMLA) aimed at applications that make use of trade-offs between different metrics. The main goal of the GMLA approach is to improve the communication efficiency, in a communication environment where the network topology is unknown to the base station.

Simulations were performed on random star topologies assuming different levels of faults. Observing Figure 5 it can be notice that the communication efficiency maintains the same level when IEEE 802.15.4 is used. However, when GMLA is used, it is possible to notice a gain of almost 10% in communication efficiency.

---

[1] freely available at http://www.control.lht.se/truetime.

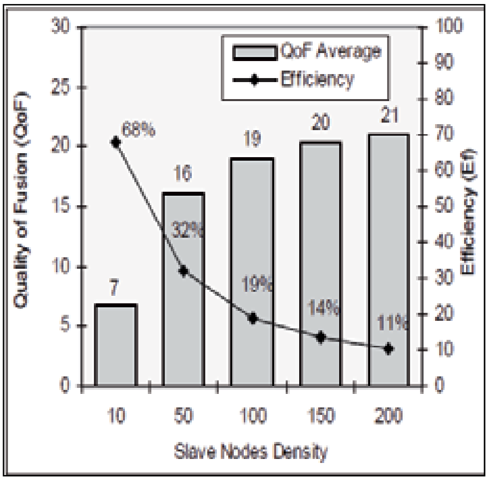


Figure 4. IEEE 802.15.4 simulated behaviour.

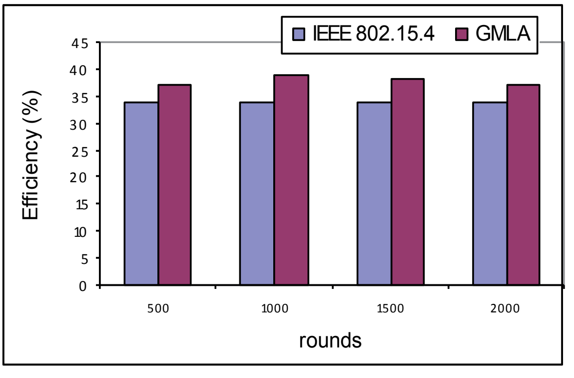


Figure 5. Comparison of GMLA and IEEE 802.15.4.

It is possible to notice that IEEE 802.15.4 presents a static behavior, and that it does not learn better communication patterns when topology changes are faced (Figure 6).

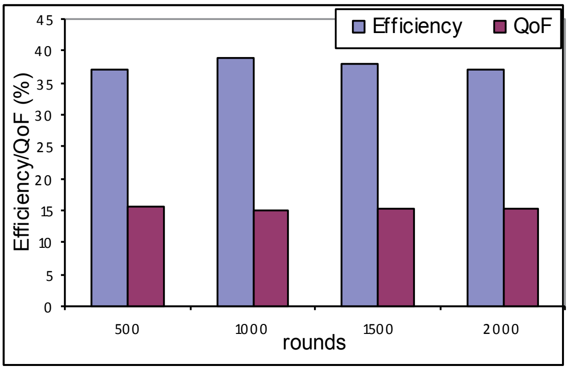


Figure 6. GMLA Efficiency and QoF values.

An analysis of Figure 7 indicates that the *QoF* is maintained at almost the same level, in all simulations. However, the higher level of *Ef* was achieved with 1,000 round simulations. This may be explained through GMLA´s learning behaviour, which tries different configurations when longer simulations are run.

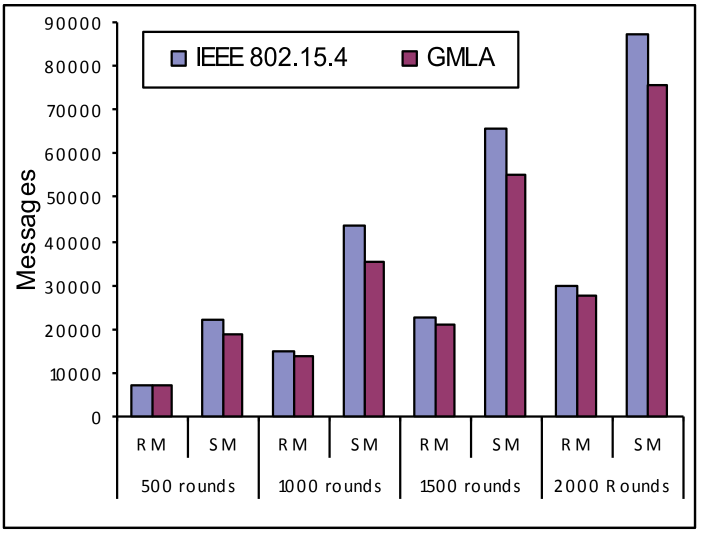


Figure 7. Sent (SM) and Received Messages (RM).

Consequently, it is possible to conclude that the GMLA approach is able to do the trade-off between *QoF* and *Ef* and GMLA uses lower levels of energy than IEEE 802.15.4. However, this approach is only suitable for applications with a homogeneous signal throughout the entire monitoring area.

Also in (Pinto e Montez, 2010), a Variable Offset Algorithm (VOA), which targets the optimization of the communication efficiency in dense WSNs with star topology, is proposed. The VOA can be easily implemented into IEEE 802.15.4 devices, as it is a light middleware that is implemented at the application layer. The main target of VOA is the communication efficiency through the use of random offsets before the transmission of data by the slave nodes.

The VOA algorithm was assessed with the help of an experimental setup based on real situations and one of the experiments has been performed by varying the number of slave nodes. The results are shown in Figure 8. The goal was to evaluate the influence of the number nodes on the *Ef* and *QoF* metrics. When compared with VOA, IEEE 802.15.4 presents similar results for just one case: a network with 4 slaves. When the number of slaves increases, the difference between VOA and IEEE 802.15.4 become greater. The difference of efficiency between VOA and IEEE 802.15.4 when considering 29 slaves is of more than 100%. These results show that VOA has a satisfactory performance and maintains a minimum QoS level even with a high number of slaves (Pinto e Montez, 2010).

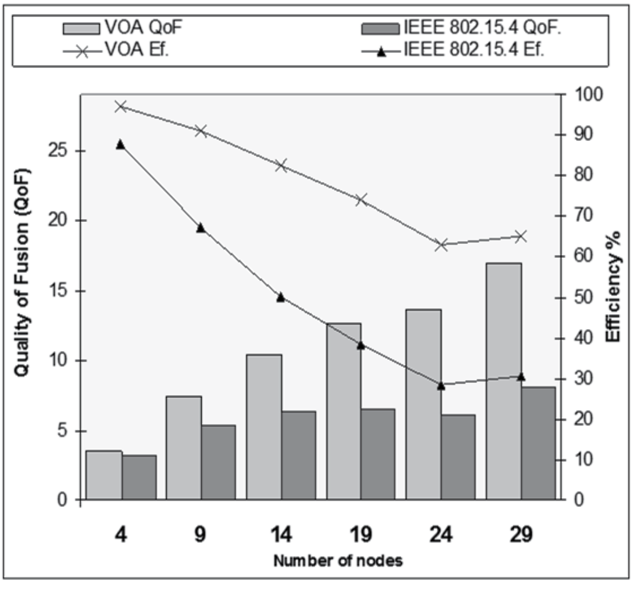


Figure 8. Comparison of VOA and IEEE 802.15.4 protocol with variable number of nodes.

The Decentralized Power-Aware Wireless Sensor Network (DPAWSN) approach has the main goal of maintaining a minimum *QoF*, while improving the *Ef* and saving energy. This approach can be considered as decentralized, due to the fact that nodes have autonomy to decide whether to send or not to send messages. On the one hand, a certain *QoF* level is imposed by network administrator, on the other hand, the WSN´s lifetime is increased by the power awareness decision taken in each node.

The main idea behind DPAWSN is that the base station will control each node in order to decreases or increases the transmission rate when the *QoF* level is above or below the target value. Thus, DPAWSN is able to maintain a *QoF* level and increase the WSN´s lifetime due to the fact that nodes present a selfish behaviour as the nodes transmission rate is calculated based on individual remaining voltage.

The test assessment was conducted in a noisy environment with an unspecified number of computers communicating using IEEE 802.11.

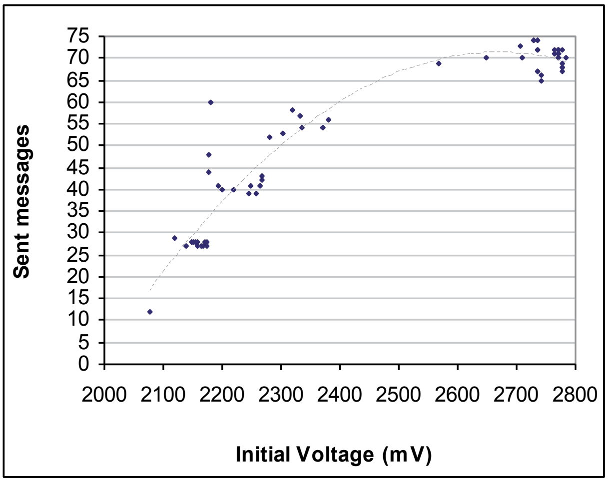


Figure 9. Figure. 9. Dispersion Graphic of 30 tests (each point represents sent messages of lower or higher battery level nodes).

Figure 9 shows the correlation between sent messages and the initial voltage in each node. Note that the initial value on the x-axis is 2000mV, which represents the minimum battery conditions for a node to work. It is possible to notice that two regions are present in the graph: the first one represents sent messages of nodes with a lower level of battery and the second one represents sent messages of nodes with a higher level of battery. In more detail, each point represents the average number of sent messages (y-axis) over the initial voltage average of the node (x-axis). The effectiveness of DPAWSN is confirmed by Figure 9, since lower battery level are applied to nodes that send less messages and nodes with higher battery level do send more messages.

Finally, DPAWSN is able to autonomically adjust the transmission rate based on the voltage level of the nodes. Moreover, it is able to achieve a targeted *QoF* imposed by system administrators.

7. Final considerations and future directions

WSNs represent one of the most interesting solutions to monitor and sense data in hazardous or inaccessible areas. This Chapter presented various possible approaches that designers might adopt in order to provide an energy efficient WSN. Due to the enormous variety of environments that these networks may be used in, numerous challenges and constraints have to be considered when choosing the optimization approach. Also, different goals and applications call for different targets that may not be achieved up to a satisfying level at the same time. Therefore, some approaches offer to define a trade-off between opposing aims, such as *QoS* and energy consumption. Each solution presented in this Chapter is aiming at solving at least one identified cause of extensive energy consumption. The approaches are using the WSN's different design levels to increase the network's lifetime, *QoS*, and/or optimize other points that may or may not suit the designer.

Taking into account that lifetime maximization is one of the main goals of WSN approaches, and as the WSN devices consume more energy during the transmission and reception of packets, even in short distances, than other tasks such as those related to processing, sensing and data storing, the design of efficient protocols and communication algorithms is a research direction.

The current tendency goes towards solutions that involve a trade-off between more than one constraint or that may adapt or change the behaviour of the WSN's nodes during its employment are showing that researches are aware of the complexity and unpredictability of the environment and task of such networks.

However, some existing research areas are becoming more relevant, mainly due to the recent technology evolution of these networks. For example, the most popular motes in the years 2000-1010 were based on 4 to 8 MHz processors and 128 Kbytes memory; but recently there are motes with 180MHz processor and up to 4Mbytes memory, supporting Java Virtual Machine. Therefore, the hardware evolution trend directs researches to implement more sophisticated and robust approaches in an autonomic and distributed way with multi-objective optimization approaches, however with power consumption as an important goal.

The gradual replacement of very expensive centralized sensor systems by a set of wireless sensor nodes, which operate in a collaborative and autonomic way (mainly with self-management and self-healing properties) is also becoming a trend. Thus, multi-agent approaches and lightweight optimization techniques are emerging as an alternative. This fact is mainly due to the distributed and optimized way that these approaches perform.

Moreover, the gradual increase in the motes´ local storage capacity has induced the use of WSN data mules. The respective research focuses on the development of architectures and algorithms where nodes must locally storage their sensed data until mobile nodes (mobile base stations) gather this information.

As final consideration it is to be stated that the variety of challenges has generated an even greater number of approaches to deal with the concerns that WSNs face in all the possible harsh and noisy environments they may face today. It is now the designers' task to find the best match or combination to optimize the network according to its environment, its tasks, and its most important requirements and constraints. As there is no solution that may cover all problems at the same time, the correct analysis of problems to be expected has become one of the most important parts of the work of today's designers.

Author details

A.R. Pinto
Universidade Estadual Paulista, UNESP, Brazil

L. Bolzani Poehls and F. Vargas
Engineering School, Catholic University of Rio Grande do Sul, PUCRS, Brazil

C. Montez
PGEAS, Universidade Federal de Santa Catarina, UFSC, Brazil

8. References

Akyildiz, I. F.; Su W.; Sankarasubramaniam Y. & Cayirci E. (2002). A Survey on Sensor Networks. *IEEE Communications Magazine,* pp. 102-114

Akyildiz, I. F.; Brunetti, F. & Blazquez, C. (2008). Nanonetworks: A new communication paradigm. *Computer Networks*, No. 52, pp. 2260–2279

Allipi, C. & Galperti, C. (2008). An Adaptive System for Optimal Solar Energy Harvesting in Wireless Sensor Networks Nodes. *IEEE Transactions on Circuits and Systems*, Vol. 55, No. 6, pp. 1742-1750

Campobello, G.; Leonardi, A. & Palazzo, S. (2012). Improving Energy Saving and Reliability in Wireless Sensor Networks Using a Simple CRT-Based Packet-Forwarding Solution. *IEEE/ACM Transactions on Networking*, Vol. 20, No. 1

Chen, W.; Neely, M. J. & Mitra, U. (2008). Energy-Efficient Transmissions With Individual Packet Delay Constraints. *IEEE Transactions on Information Theory*, Vol. 54, No. 5.

Chou, P.H. & Park, C. (2005). Energy Efficient Platform Designs for Real-World Wireless Sensing Application. *Proceedings of IEE/ACM International Conference Computer-Aided Design*

Costa, A. & Sayeed, A. (2003). Data Versus Decision Fusion for Distributed Classification in Sensor Networks. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 832-835

Demirkol, I.; Ersoy, C. & Alagöz, F. (2006). MAC Protocols for Wireless Sensor Networks: A Survey, *IEEE Communications Magazine*

Deying, L.; Xiaohua, J. & Hai, L. (2004). Energy Efficient Broadcast Routing in Static Ad Hoc Wireless Networks. *IEEE Transactions on Mobile Computing*, Vol. 3, No. 2

Felemban, E.; Lee, C. & Ekicin, E. (2006). MMSPEED: Multipath Multi-SPEED Protocol for QoS Guarantee of Reliability and Timeliness in Wireless Sensor Networks. *IEEE Transactions on Mobile Computing*, Vol.5, No. 6, pp. 738-754

Gholamzadeh, B. & Nabovati, H. (2008). Concepts for Designing Low Power Wireless Sensor Networks, *World Academy of Science, Engineering and Technology*

Guha, S.; Basu, P.; Chau, C. & Gibbens, R. (2011). Green Wave Sleep Scheduling: Optimizing Latency and Throughput in Duty Cycling Wireless Networks. *IEEE Journal on Selected Areas in Communications*, Vol. 29, No. 8

Hadim, S. & Mohamed, N. (2006). Middleware Challenges and Approaches for Wireless Sensor Networks. *IEEE Distributed Systems Online*, Vol. 7, No. 3

Ho I. & Liew S. (2007). Impact of power control on performance of IEEE 802.11 wireless networks. *IEEE Trans. Mobile Comput.*, Vol. 6, No. 11, pp. 1245-1258, November 2007

Hong, J.; Jang, I.; Lee, H.; Yang, S. & Yoon, H. (2010). MRMAC: Medium Reservation MAC Protocol for Reducing End-to-End Delay and Energy Consumption in Wireless Sensor Networks. *IEEE Communications Letters*, Vol. 14, No. 7

Huebscher, M. C. & McCann, J. A. (2008). A survey of autonomic computing degrees, models, and applications. *ACM Comput Surveys*

IEEE. (2008) Especificação IEEE 802.15.4, Available from : http://www.ieee802.org/15/pub/TG4.html

Ilyas, M. & Mahgoub, I. (2005). *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, CRC Press Inc., Boca Raton, USA

Jungmin, S.; Jintae, K. & Indanil, G. (2005). Cushion: Autonomically Adaptative Data Fusion in Wireless Sensor Networks, *Proceedings of MASS'05*

Karl, H. & Willig, A. (2005). *Protocols and Architectures for Wireless Sensor Networks*, John Willey & Sons

Kephart, J.O. , Chess, D.M., The Vision of Autonomic Computing, IEEE Computer, 2003.

Kim, S.; Pakzad, S.; Culler, D.; Demmel, J.; Fenves, G.; Glaser, S. & Turon, M. (2007) Health Monitoring of Civil Infrastructures Using Wireless Sensor Networks. *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, Massachusetts, USA

Kim, J. & Know, Y. (2008). Interference-Aware Transmission Power Control for Wireless Sensor Networks. *IEICE Trans. Commun.*, Vol. E91-B, No. 11, November 2008

Kim, M.; Han, Y. & Park, H. (2011). Energy-Aware Hybrid Data Aggregation Mechanism Considering the Energy Hole Problem in Asynchronous MAC-Based WSNs. *IEEE Communications Letters*, Vol. 15, No. 11, November 2011

Koubaa, A.; Severino, R.; Alves, M. & Tovar, E. (2009). Improving Quality-of-Service in Wireless Sensor Networks by mitigating "Hidden-Node Collisions". *IEEE Transactions on Industrial Informatics*, Special Issue on Real-Time and Embedded Networked Systems, Vol. 5, No. 3, August 2009

Kubisch, M.; Karl, H.; Wolisz, A.; Zhong, L. C. & Rabaey, J. (2003). Distributed Algorithms for Transmission Power Control in Wireless Sensor Networks, *Proceedings of WCNS03*, IEEE Wireless Communications and Networking Conference, March 2003

Lavratti, F.; Ceratti, A., Prestes, D.; Bolzani, L.; Vargas, F.; Montez, C.; Hernandez, F. & Gatti, E. (2012). A Transmission Power Self-Optimization Technique for Wireless Sensor Networks. *ISRN Communications and Networking*.

Lee, W. L.; Datta, A. & Cardell-Oliver, R. (2008). FlexiTP: A Flexible-Schedule-Based TDMA Protocol for Fault-Tolerant and Energy-Efficient Wireless Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 19, No. 6

Lin, S.; Zhang, J.; Zhou, G.; Gu, L.; He, T. & Stankovic, J. A. (2006). ATPC: Adaptive Transmission Power Control for Wireless Sensor Networks. *Proceedings of SenSys'06*, November 2006

Melodia, T.; Pompili, D. & Akyildiz, I.F. (2010). Handling Mobility in Wireless Sensor and Actor Networks. *IEEE Transactions on Mobile Computing*, Vol. 9, No. 2, pp. 160-173

MicaZ. Mote Datasheets, 15/03/2012, Available from: www.xbow.com

Molla, M. & Ahamed, S. I., A Survey for Sensor Network and Challenges. (2006). *Proceedings of the 2006 International Conference on Parallel Processing Workshops, 2006*

Monks, J. P.; Bharghawan, W.; Mei, W. & Hwu, W. (2001). Power Controlled Multiple Access Protocol for Wireless Packet Networks. *Proceedings of INFOCOM*, pp. 219-228

Morita, K.; Watanabe, K.; Hayashibara, N.; Enokido, T. & Takanizawa, M. (2007). Efficient Data Transmission in a Loss and Resource Limited Wireless Sensor-Actuator Network. *Proceedings of the 10th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing* , 2007

Oliver, R. S. & Fohler, G. (2010). Timeliness in Wireless Sensor Networks: Common Misconceptions. *Proceedings of the 9th International Workshop on Real-Time Networks*, Brussels, Belgium, July 2010

Omeni, O.; Wong, A. C. W.; Burdett, A. J. & Toumazou, C. (2008). Energy Efficient Medium Access Protocol for Wireless Medical Body Area Sensor Networks. *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 2, No. 4, December 2008

Otal, B.; Alonso, L. & Verikoukis, C. (2009). Highly Reliable Energy-Saving MAC for Wireless Body Sensor Networks in Healthcare Systems. *IEEE Journal on Selected Areas in Communications*, Vol. 27, No. 4

Patil, S.; Das, S. K. & Nasipuri, A. (2004). Serial Data Fusion Using Space-filling Curves in Wireless Sensor Networks. *Proceeding of IEEE SECON*, pp. 182-190, 2004

Pinto, A.R., Montez, C. Autonomic Approaches for Enhancing Communication QoS in Dense Wireless Sensor Networks with Real Time Requirements, 2010 IEEE Int. Test Conference, 2010. 1-10.

Quan, Z.; Subramanian, A. & Sayed, A. H. (2007). REACA: An Efficient Protocol Architecture for Large Scale Sensor Networks. *IEEE Transactions on Wireless Communications*, Vol. 6, No. 10, pp. 3846-3855, 2007

Raghunathan, W.; Gareniwal, S. & Srivastava, B. (2006). Emerging Techniques for Long Lived Wireless Sensor Networks. *IEEE Communications Magazine*, Vol. 44, No. 4, pp. 108-114, April 2006.

Rossi, M. & Zorzi, M. (2007). Integrated Cost-Based MAC and Routing Techniques for Hop Count Forwarding in Wireless Sensor Networks, IEEE Transactions on Mobile Computing, Vol. 6, No. 4, April 2007

Shi, L. & Fapojuwo, A. O. (2010). TDMA Scheduling with Optimized Energy Efficiency and Minimum Delay in Clustered Wireless Sensor Networks. *IEEE Transactions on Mobile Computing*, Vol. 9, No. 7, July 2010

Sohraby, K.; Minoli, D. & Znati, T. (2007). *Wireless Sensor Networks: Technology, Protocols and Applications*, John Wiley & Sons

Stankovic, J.A.; Abdelzaher, T.F.; Lu, C.; Sha, L. & Hou, J. C. (2003). Real-Time Communication and Coordination in Embedded Sensor Networks. *Proceedings of the IEEE*, Vol. 91, No. 7, pp. 1002-1022, 2003

Tseng, H.; Sheu, S. & Shih, Y. (2011). Rotational Listening Strategy for IEEE 802.15.4 Wireless Body Networks. *IEEE Sensors Journal*, Vol. 11, No. 9

Wu, Q.; Rao, N. S. V.; Barhen, J.; Iyenger, S. S.; Vaishnavi, V. K.; Qi, H. & Chakrabarty, K. (2004). On Computing Mobile Agent Routes for Data Fusion in Distributed Sensor Networks. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 6

Wu, X.; Chen, G. & Das, S. K. (2008). Avoiding Energy Holes in Wireless Sensor Networks with Nonuniform Node Distribution. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 19, No. 5

Wu, Y.; Fahmy, S. & Shroff, N. (2009). Optimal Sleep/Wake Scheduling for Time-Synchronized Sensor Networks With QoS Guarantees. *IEEE/ACM Transactions on Networking*, Vol. 17, No. 5, pp. 1508- 1521

Wu, Y.; Li, X.; Liu, Y. & Lou, W. (2010). Energy-Efficient Wake-Up Scheduling for Data Collection and Aggregation. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 21, No. 2

Yick, J.; Mukherjee, B. & Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks*, no. 52, pp. 2292–2330

Zhao, L.; Guo, L.; Zhang, J. & Zhang, H. (2009). Game-theoretic medium access control protocol for wireless sensor networks. *IET Communication*, Vol. 3, No. 8, pp. 1274–1283

Chapter 3

Radio-Triggered Power Management in Wireless Sensor Networks

Jie Wang, Qinghua Gao, Yan Yu, Hongyu Wang and Minglu Jin

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/39189

1. Introduction

As an emerging technique, the wireless sensor networks (WSNs) provide a cheap and straightforward way to acquire sensor data in hard accessible places or on rotating devices. Generally, the sensor nodes are self-supportive by including sensing, processing, storage and communication capabilities. And the node hardware usually operates with batteries, as it is infeasible to sustain the nodes with power supply, and other ambient sources of energy (like solar cells and wind energy) do not provide the required power levels, require the specific operational conditions or negatively impact from the shape. In most of the WSNs applications, it is inconvenient or even infeasible to replace the batteries of the node, and the lifetime of the batteries is equivalent to the lifetime of the WSNs. Hence, to prolong the lifetime of the WSNs, it is an important issue to design a power management scheme for employing the wake-up/sleep schedules, so as to perform effective power management for the WSNs nodes.

The essence function of the power management scheme is to determine at what time a node should enter a high-power wake-up running mode or enter a low-power sleep mode. And an excellent power management scheme should guarantee that a node should stay in the sleep mode as more as possible, since the power consumption of the two modes is dramatically different. For example, if the node adopts the MSP430 processor as the Micro Control Union (MCU), which has five power saving modes with different energy saving features. In one of the working modes, the processor shuts down all the components except for the memory, and the current consumption is only 0.1uA. This is less than 0.1% of the active working mode. So the lifetime of the node can be remarkably increased if it stays in the low-power sleep mode as long as possible. The simplicity of changing the wake-up mode to the sleep mode by applying a set of MCU instructions that shut down the hardware components is universally acknowledged. However, difficulties exist in the design process of changing the sleep mode to the running mode when the MCU of the WSNs node is in deep sleep mode and is unconscious of when to wake up. However, the messages can only be received and processed when the node is in the wake-up running mode. Hence, how to develop a low cost, low power consumption scheme to wake up the node from the sleep mode on demand becomes a challenging and promising task.

To solve this problem, many power management schemes have been proposed. The most widely adopted scheme is to realize wake-up/sleep schedule by waking up the WSNs nodes periodically. However, it is hard to define the wake up period, as the energy would be wasted if the node wakes up too often, and the network controller's commands would be missed if the node wakes up too seldom. A common phenomenon is that nothing happened in most of the wake-up periods, and the node enters sleep mode again. The disadvantage of this scheme is that it can not ensure that the node stays in the running mode when needed, meanwhile, the periodically wake up operation consumes excessive energy, which makes the scheme a high power consumption method. Another method is to use a low power stand-alone radio receiver subsystem to monitor the environment [7, 10]. The node keeps sleeping unless the receiver wakes it up after receiving the wake-up signal. For example, Atmel's ATA5283 is an ideal chip to do this work [7]. The newly proposed PicoRadio has such function too [10]. However, the consumption of extra energy and cost is the negative part of this method.

In this chapter, we present a novel concept for designing a passive radio-triggered wake-up circuit for realizing the wake-up/sleep schedules, which is made up of some diodes, capacitors, and inductors, and could transform the radio frequency (RF) power to direct current (DC) voltage efficiently. Similar to the stand-alone radio receiver wake-up scheme, the activation and shutting down of the WSNs node is controlled by detecting whether there is a wake-up signal in the environment, and thus eliminates the energy wasting wake-up periods. When the network controller wants to wake up the node, it can emit the wake-up signal, the passive radio-triggered wake-up circuit harvests the RF power and produces a DC voltage to trigger the interrupt of the MCU, so as to wake up the sensor node. However, unlike the stand-alone radio receiver which consumes extra energy and cost, the radio-triggered wake-up circuit is made up of low cost diodes, capacitors, and inductors, and the circuit itself is a passive circuit which does not need any power supply. The above advantages make the passive radio-triggered wake-up circuit an ideal choice for realizing the power management in the cost sensitive WSNs.

The chapter is organized as follows. The following section gives the background and related work. Section 3 describes the system view of the radio-triggered wake-up circuit. Section 4 presents the detailed circuit level design. Section 5 provides the simulation and measurement results. Section 6 discusses some schemes to improve the performance of the basic circuit, such as the schemes to improve the wake-up distance and to realize addressable wake-up. And the conclusion is drawn in the section 7.

2. Related work

The essential function of the power management scheme is to reduce the power consumption of the node consumed by unused hardware peripherals and radio resource. While it is easy to define when transmission is required, reception is usually unpredictable and asynchronous to a sensor node. However, the messages can only be received and processed when the node is in the wake-up running mode. Hence, how to develop a low cost, low power consumption scheme to wake-up the node from the sleep mode on demand becomes a challenging and promising task.

The simplest power management schemes are continuous listening or duty-cycling. Besides having energy conservation, an evident drawback of the duty cycling schemes is the increased latency compared to the always-on mode, and the scheme can not ensure that the node stays in the running mode when needed. Some advanced energy-efficient multi-access control (MAC) protocols like WiseMAC protocol and SCP-MAC protocol have been proposed, the detailed protocols can be referred to the survey of WSNs specific MAC protocols [1, 12, 15]. The wake-up costs may be amortised over periods of inactivity, in which MAC protocols regularly perform carrier sense operations. A work [8] explores several schemes for predicting the arrival time of the next message after an initial wake-up, which allows for scheduling an efficient carrier sense. More recently, a radio-triggered MAC protocol was proposed to realize power management [13].

One of the possible alternatives in minimizing latency is to use an additional wake-up radio hardware, which is thoroughly optimized for negligible power consumption and is capable to react instantly on an event of interest. Atmel's ATA5283 is an ideal chip for power management. It is an ultra-low power receiver works at 125KHz. The power consumption of the chip is 10uW with the supply voltage of 2V. However, the chip can only work in the ultra-low frequency of 125KHz, so a large scale antenna is required to receive the signal, which restricts the chip from adopting in lots of applications. PicoRadio [10] uses a dedicated designed low power transceiver module (build as a prototype IC), which is capable of monitoring radio environment. It can be used as a stand-alone radio module on the sensor node to receive the wake-up signal emitted by the network controller. The total power consumption of the PicoRadio module in the receive mode is 380uW with the receiver sensitivity of -75dBm and supply voltage of 1V. Although the power consumption is much less than many state-of-the-art normal radio transceiver, which typically use 2-3 V supply voltage and consumes 20-90mW, this solution still consumes significantly high power consumption while staying always-on.

Some literatures [3, 5, 9, 11, 14, 17] present the design of the power harvesting circuit for the radio frequency identification devices (RFID) chip. They adopt the high efficiency rectifier to realize the RF to DC signal transformation. They designed the circuit as a module for the RFID IC chip. To design a chip to realize the wake-up function is too complex and expensive for the WSNs node. Hence, we realize the design using discrete components. It's cheaper and more flexible and can be applied to different applications with little change. Our proposed passive radio-triggered wake-up circuit [16] is made up of some diodes, capacitors, and inductors, and could transform the RF power to DC voltage efficiently, and the circuit itself is a passive circuit which does not need any power supply. The above advantages make the passive radio-triggered wake-up circuit an ideal choice for realizing the power management in the cost sensitive WSNs.

3. System overview

The WSNs node with passive radio-triggered wake-up circuit can be waked up by means of harvesting power from RF signal emitted by the network controller. And the network controller must transmit a high intensity RF signal through the air. When it is the time for the node to be waked up, the controller emits the wake-up signal, and the circuit harvests the RF signal and produces a DC interrupt signal to wake up the MCU of the node immediately. Fig. 1 illustrates the system view of the network controller and node. The whole WSNs is made up of the general node equip with the passive radio-triggered wake-up circuit, and the network controller which could transmit high intensity RF wake-up signal. It should be mentioned that since our proposed passive radio-triggered wake-up circuit is made up of discrete devices, it is very convenient to change the working frequency of the circuit. Hence, most of the commercial transmitters can be adopted to realize the network controller design. In the WSNs applications, one can integrate the network controller's function into the WSNs node for simplicity.

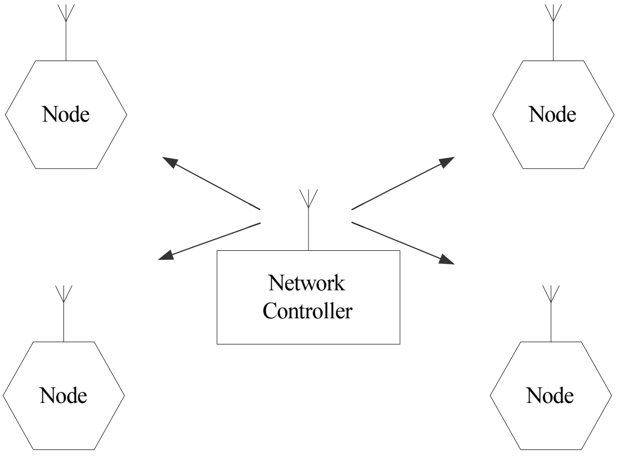


Figure 1. Illustration of the network controller and node system

The block diagram of the WSNs node is shown in the Fig. 2. Besides the traditional modules of the node, a passive radio-triggered wake-up circuit which transforms the RF energy to DC voltage is added to the node. It shares the same antenna with the transceiver, and sends the DC voltage directly to the MCU to trigger the interrupt operation.

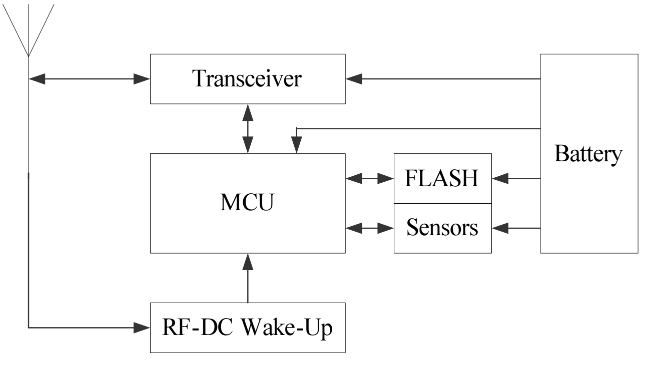


Figure 2. Block diagram of the node equip with the radio-triggered wake-up circuit

In this chapter, the RF to DC passive radio-triggered wake-up circuit is designed to operate in the UHF frequency ISM band of 915MHz, it should be mentioned that it is quit convenient to change the working frequency of the circuit. The modules of the RF to DC radio-triggered wake-up circuit are shown in Fig. 3. The circuit is made up of an antenna, an impedance matching network, and a rectifier. The antenna picks up the RF power sent out by the network controller, the impedance matching network ensures maximum power transfer in the system, and the rectifier converts the RF power to a DC voltage. In the Fig. 3, the signal at the mark 1 is RF signal, the signal at the mark 2 is also RF signal, however, the signal at the mark 3 is DC voltage. The rectifier realizes the crucial RF to DC transformation task, and the antenna and impedance matching network are auxiliary circuits that improve the RF to DC transformation efficiency.

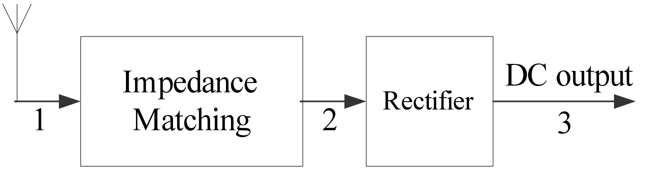


Figure 3. RF to DC radio-triggered wake-up circuit

The RF power received by the wake-up circuit at the mark 1 of Fig. 3 is given by the following equation

$$P_r = \frac{P_s G_s G_r \lambda^2}{(4\pi d)^2} \tag{1}$$

where P_s is the transmission power of the network controller, G_s is the gain of the antenna of the network controller, G_r is the gain of the node antenna, λ is the wavelength of the electromagnetic wave, and d is the distance between the network controller and the node. In order to increase the working distance of the network controller, one can increase the transmission power of the network controller or increase the antenna gain of network controller or node. The transmission power P_s is 4W for our prototype 915MHz design. We assume that the antenna gain is 1dB. The wavelength λ of 915MHz is 0.328m. If the distance between the network controller and the node is 40m, the received power P_r will be 1.7uW (-27.7dBm).

The impedance matching network transforms the input impedance of the rectifier to that of the antenna. A good impedance matching between the antenna and rectifier can reduce transmission loss and increase voltage gain [2]. Another function of the impedance matching network is to passively amplify the voltage. The theoretical amplifier efficiency of a matching network is expressed by the following equation

$$\frac{V_{out}}{V_{in}} = \frac{1}{2}\sqrt{1+Q^2} \tag{2}$$

where V_{out} is the output voltage of the matching network (the voltage of mark 2 in Fig. 3), V_{in} is the input voltage of the matching network (the voltage of mark 1 in Fig. 3), and Q is the quality factor of the matching network. Equation (2) implies that the maximum gain of the matching circuit can be achieved with the maximum quality factor Q. However, one drawback of the high Q circuit is that it reduces the bandwidth of the circuit. The relation of the quality factor Q and bandwidth Δf can be expressed as follows

$$Q = \frac{f_c}{\Delta f} \tag{3}$$

where f_c is the center frequency of the wireless signal, and Δf is the bandwidth of the circuit.

The rectifier is the most important module of the RF to DC circuit. It converts the RF energy to DC voltage. Because the peak voltage of the RF signal at the input of the rectifier is very low, we adopt the multistage rectifier in the design. Some papers have studied the rectifier design for passive RFID tags [14, 17]. However, all of them use CMOS process to design a chip to realize the function. Adding a chip to realize the wake-up function is too complex and expensive for the WSNs node. Hence, we realize the design using discrete components. It's cheaper and more flexible and can be applied to different applications with little change. Details of the design will be given in section 4.

4. Circuits design

4.1. Basic voltage doubler rectifier

The basic voltage doubler rectifier is shown in Fig. 4. It is made up of two diodes and two capacitors. The two diodes are connected in series, oriented so that forward current can only flow from the ground potential to the positive terminal of the output voltage V_{out}. The capacitor C1 prevents the DC current from flowing into the circuit. It stores the charge and permits the high frequency currents to flow. The capacitor C2 stores the resulting charge to smooth the output voltage V_{out}. Essentially, the circuit is a charge-pump structure. The capacitor C1 and diode D1 make up a dc-level shifter, and the capacitor C2 and diode D2 form a peak detector.

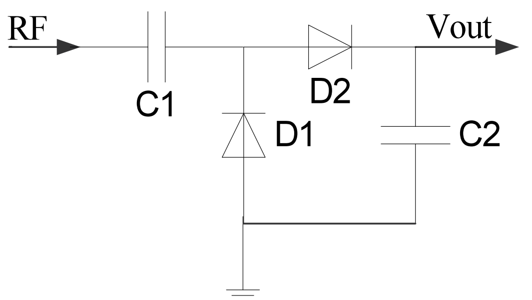


Figure 4. Basic voltage doubler rectifier

When the RF signal is negative and larger than the turn-on voltage of the diode D1, the current flows from the ground point through the diode D1 and causes charge to accumulate on the capacitor C1. The equivalent model of the negative part of the RF cycle is shown in Fig. 5. At the negative peak, the voltage across the capacitor C1 is the difference between the negative peak voltage V_{pk} and the voltage on top of the diode V_{on}. The right (positive) plate of the capacitor will be charged up to the voltage of $V_{pk} - V_{on}$. As a result, the DC level of the signal applied to the peak detector (capacitor C2 and diode D2) is shifted to the voltage V_{shift} as

$$V_{shift} = V_{pk} - V_{on} \tag{4}$$

When the RF signal is positive and larger than the turn-on voltage of the diode D2, the diode D1 turns off and the diode D2 turns on. The current flows from the input capacitor C1 through the output diode D2 to the output capacitor C2. The equivalent model of the positive part of the RF cycle is shown in Fig. 6. The peak voltage over diode D2 can be calculated by

$$V_{out} = V_{shift} + V_{pk} - V_{on} = 2\left(V_{pk} - V_{on}\right) \tag{5}$$

The output voltage V_{out} is defined by adding the voltage across the capacitor C1 to the peak positive RF voltage, and subtracting the turn-on voltage V_{on}. When the received RF power

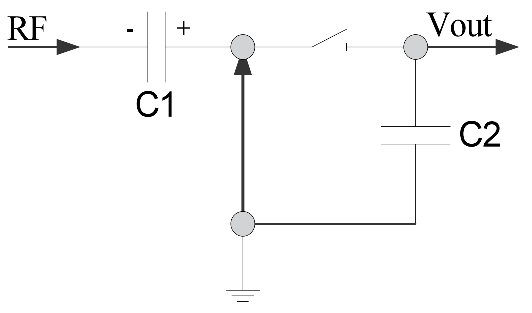


Figure 5. Rectifier during the negative part of the RF cycle

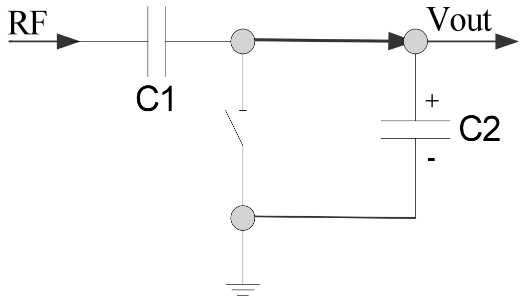


Figure 6. Rectifier during the positive part of the RF cycle

is invariable, in order to increase the output voltage V_{out}, the voltage on top of the diode V_{on} must be as low as possible. Hence, the zero bias schottky diode is a good choice for the rectifier design. In theory, the output voltage V_{out} could be double the peak voltage of the RF signal when the turn-on voltage V_{on} can be ignored, from which fact the circuit derives its name. In the real application, the load resistance of the capacitors can not be ignored, and the circuit can not be able to draw out all the current and charge that stored in the capacitors, hence, the real output voltage V_{out} must be less than the theory value.

4.2. Multistage rectifier

One stage rectifier can only produce a voltage of $2\left(V_{pk} - V_{on}\right)$. In most of the scenes, the voltage is not high enough to trigger the MCU, hence, multistage rectifier is proposed.

As shown in Fig. 4. If we changed the ground level to a DC voltage V_{ref}, as has been analyzed above, the output voltage V_{out} will be $2\left(V_{pk} - V_{on}\right) + V_{ref}$. So we can cascade the basic voltage doubler rectifiers to form the multistage rectifier. The circuit diagram of the multistage rectifier is shown in Fig. 7. We can see that the multistage rectifier is made up of several one stage rectifiers, and the DC voltage generated at each stage is applied as the DC reference to its following stage. Therefore, the DC output voltage of the N-stage multistage rectifier is expressed as follows

$$V_{out} = 2N \times (V_{pk} - V_{on}) \tag{6}$$

Equation (6) implies that one could continue to add as many stages as required to convert even the most modest input RF voltage into a proper output voltage V_{out}. But as stages are added, more and more power and charge are wasted in the diodes, capacitors, and PCB lines. The more the stages are added, the less efficient the circuit will be. What we want is the maximum of the output voltage as far as the radio-triggered wake-up function is concerned. Hence, the number of the stages must be carefully designed.

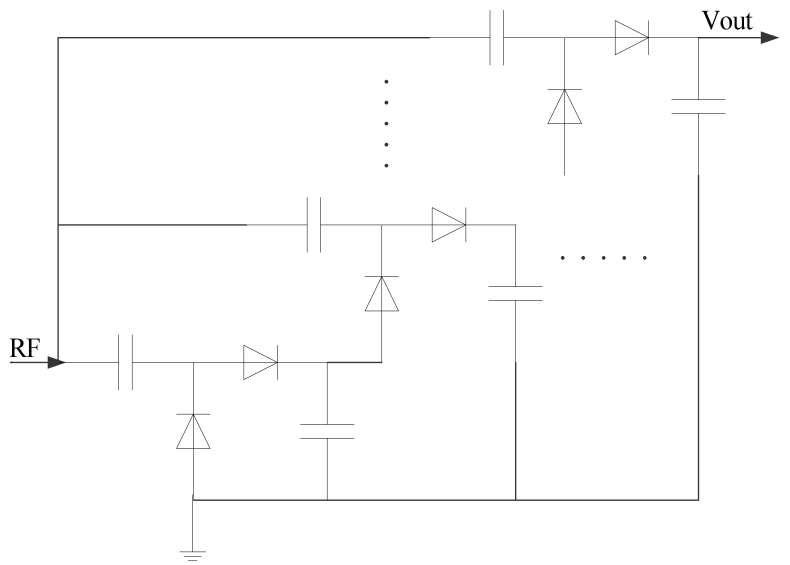


Figure 7. Multistage rectifier

We adopt Avago's zero bias schottky diode HSMS2852 in our prototype design. At a small signal level, the schottky diode can be represented by a linear equivalent circuit, as shown in Fig. 8. Where Rj is the junction resistance of the diode, Cj is the junction capacitance of the diode, Rs is the parasitic resistance representing the losses in the diode's bond wire and the bulk silicon at the base of the chip, and Lp and Cp are parasitic inductance and capacitance caused by the package. As far as the HSMS2852 with SOT-23 package which is used in the prototype design is concerned, Rj is 9Kohm, Cj is 0.16pF, Rs is 20ohm, Cp is 0.08pF, and Lp is 2nH. We utilize the ADS software from Agilent to simulate the input impedance of the multistage rectifier with different stages by using the equivalent circuit in Fig. 8. The capacitors in the multistage rectifier are set to 100pF. The simulation results are summarized in the Table 1. From the Table we can see that the resistance and reactance reduce with the increasing of stages. Hence, with the increase of stages, it becomes more difficult to design the impedance matching circuit. Meanwhile, we should keep in mind that if we use different stages of multistage rectifier, the impedance matching circuit should be redesigned to ensure the maximum power transformation.

4.3. Impedance matching network

From the Table 1 we can see that the input equivalent impedance of different stages consist of resistance and capacitive reactance. The impedance of the antenna is assumed as 50ohm for the 1/4 wavelength printed monopole antenna which is used in the design. As has mentioned above, the impedance matching network must have a high Q so as to produce high voltage for the rectifier. Hence, a simple two elements L type matching is unsuitable as the Q value of

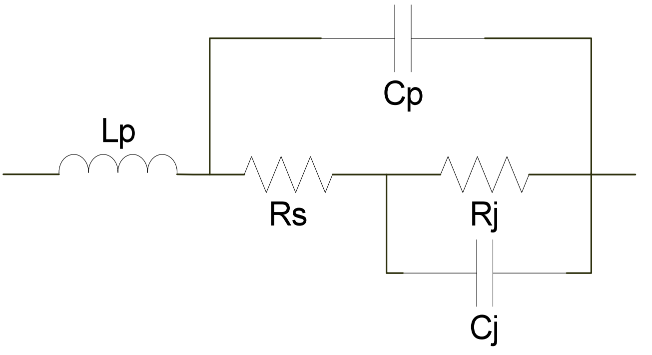


Figure 8. Equivalent circuit of the HSMS2852

Stages	Zin(ohm)	Stages	Zin(ohm)
1	30.661-j329.6	6	5.102-j55.102
2	15.311-j165.071	7	4.376-j47.235
3	10.212-j110.135	8	3.826-j41.333
4	7.655-j82.623	9	3.402-j36.743
5	6.127-j66.114	10	3.064-j33.070

Table 1. Input equivalent impedance of the multistage rectifier with different stages

the L type matching network is fixed. The three elements π type matching network is flexible to adjust the Q value as requested. So a more flexible π type matching network is adopted here. To illustrate the character of the basic one stage rectifier, a L type and a π type matching network are designed and compared. Fig. 9 shows the result of the L type matching network and Fig. 10 shows the result of the π type matching network.

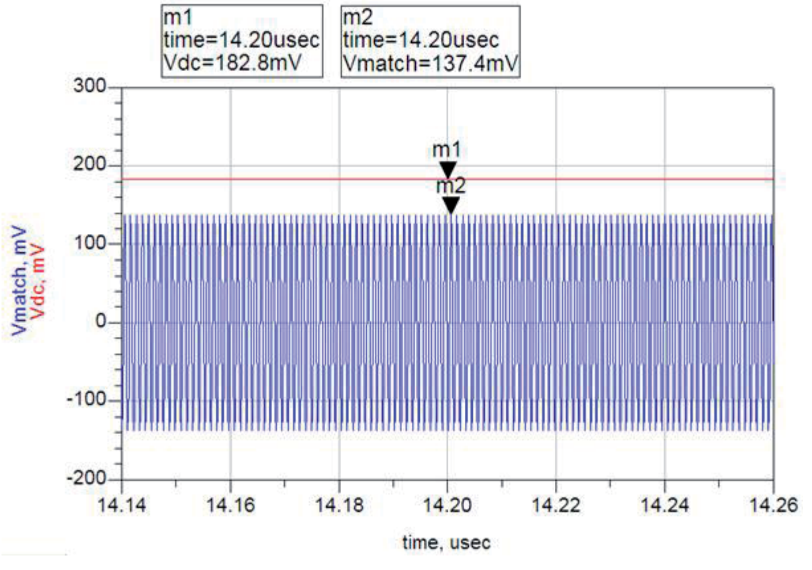


Figure 9. Voltage of L type matching network with input signal of 1.7uW (-27.7dBm)

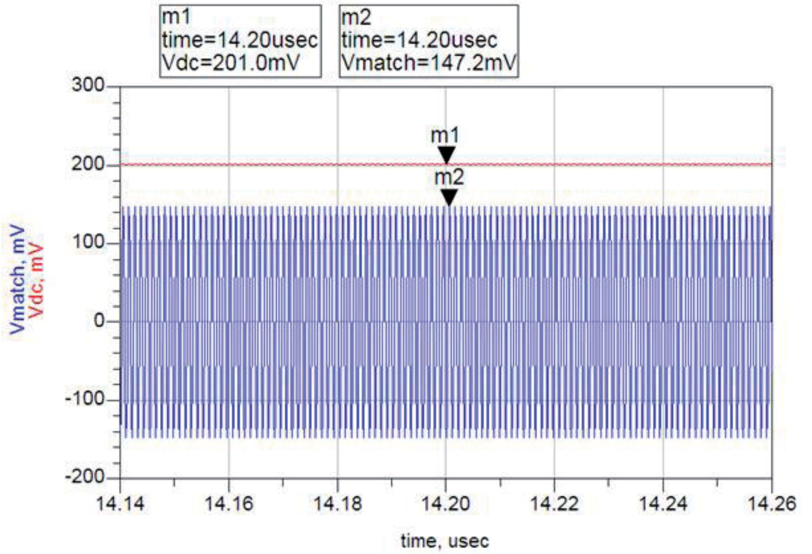


Figure 10. Voltage of π type matching network with input signal of 1.7uW (-27.7dBm)

In the Fig. 9 and Fig. 10, the Vmatch is the voltage measured at the output of matching network, and Vdc is the DC voltage output at the output of the rectifier. In the simulation, the power of the input RF signal is 1.7uW (-27.7dBm). From the figures we can see that the three elements π type impedance matching network produces a voltage higher than the two elements L type network, and this is because the π type matching network has a higher Q than the L type network. In order to produce high voltage output, it is better to use the π type matching network. As has been calculated with the equation (1), with a distance of 40m and transmitting power of 4W, the received signal strength is 1.7uW(-27.7dBm). Hence, the basic one stage rectifier circuit with good impedance matching will produce a DC voltage of about 200mV.

4.4. Optimization of the circuit

As has been discussed above, both the impedance and efficiency of the rectifier are changed with different stages. From equation (6) we can know that for a larger N a higher V_{out} can be derived. However, the optimal number of stages should be found with a compromise between high DC output and low power loss due to power consumption of the schottky diodes. Meanwhile, the V_{on} voltage of the zero bias schottky diode can not be neglected when the input signal is very weak. Hence, the more the stages, the more voltage will drop on the diodes. Table 2 shows the DC output voltage which is changed with the number of stages. The result is acquired with the input RF signal power of 100uW (-10dBm), 20uW (-17dBm), 6.8uW (-21.7dBm), and 1.7uW (-27.7dBm). As the cost is a key factor of the WSNs node design, only three stages are concerned here.

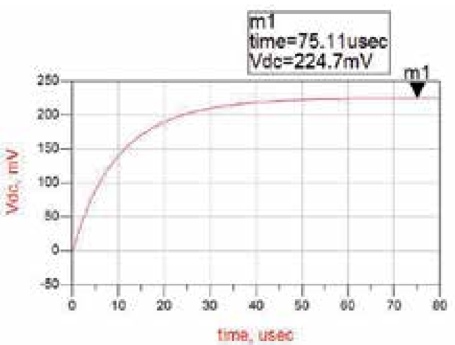
From Table 2 we can see that when the input power is high, the output voltage increases with the increase of the circuit stages obviously. However, when the input power is small, the output voltage increases with the circuit stages slightly. For the radio-triggered wake-up function, the input power is always small. Hence, two stages multistage rectifier circuit is

	Input power			
Stages	-10dBm	-17dBm	-21.7dBm	-27.7dBm
1	1.409V	813.1mV	493.5mV	200.8mV
2	2.212V	1.233V	682.7mV	224.7mV
3	2.807V	1.491V	692.7mV	225.6mV

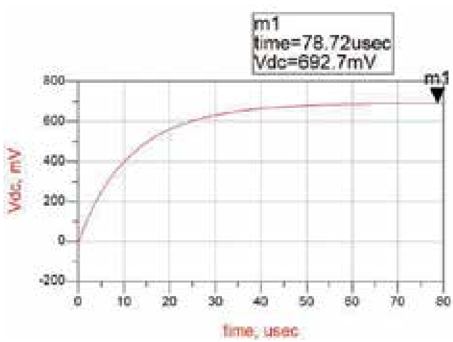
Table 2. Voltage under different input power with different stages

suitable for the radio-triggered wake-up function. However, as far as other functions such as supplying power to other IC is concerned, it is a better choice to use a multistage rectifier with more stages. We simulate the character with different number of stages, and find that a 10 stages multistage rectifier can produce a voltage of 5.016V with the input signal power of -10dBm.

The rectifier output voltage is changed with the changing of the load impedance. For larger load impedance, a higher V_{out} voltage can be derived. The simulated and measured results in this chapter are achieved with the load impedance of 100Kohm which is the general input impedance of the state-of-the-art MCU interrupt pin. With bigger load impedance, a higher voltage can be achieved.



(a) Input power of -27.7dBm



(b) Input power of -21.7dBm

(c) Input power of -17dBm

(d) Input power of -10dBm

Figure 11. Output voltage of two stages rectifier with different input power

The time domain simulation waveforms of the two stages rectifier are shown in Fig. 11. We can see that with the increasing of the input power, it will take shorter time for the output voltage to become stable. With the input power of -27.7dBm, it takes about 25us to produce a voltage of 200mV. However, with the input power of -10dBm, it takes only about 3us to produce a voltage of 200mV. In the radio-triggered wake-up application, a wake-up time of less than 30us is an acceptable time to trigger the MCU to wake up.

5. Experimental results

To evaluate the above theory and simulation analysis, we design a prototype circuit as illustrated in the Fig. 12. The prototype circuit is a 915MHz passive radio-triggered wake-up circuit which is made up of a π type impedance matching network and a two stages rectifier. For the sake of test and measurement, we use the SMA connectors to connect the circuit with the antenna, vector network analyzer, and oscilloscope. The default transmitting power of the network controller is 4W in the evaluations.

Figure 12. The prototype design

To evaluate the performance of the π type impedance matching network, we use the E5071C vector network analyzer from the Agilent to measure the return loss (S11 parameter) of the circuit. We scan the frequency band of 900MHz to 930MHz, and the measurement result is shown in Fig. 13. We can see that the return loss of the circuit is -16.9dB at the frequency of 915MHz, which is the minimum value in the whole frequency band. Hence, the circuit ensures that most of the energy can be transmitted to the rectifier circuit in 915MHz, and the π type impedance matching network achieves good performance.

We use an oscilloscope to measure the DC output voltage under different input power, and the measurement results are illustrated in the Fig. 14. We measure the DC output voltage with an input power of -27.7dBm and -21.7dBm. From Fig. 14 we can see that the measured voltage level is close to the simulation result, and the rising time of the voltage is similar to the simulation result too. These results confirm the excellent performance of the two stages rectifier circuit.

To evaluate the overall performance of the passive radio-triggered wake-up circuit, we measure and simulate the output voltage of the circuit with different input power, and plot the results in the Fig. 15. We can see that the measured results are near to the simulation results.

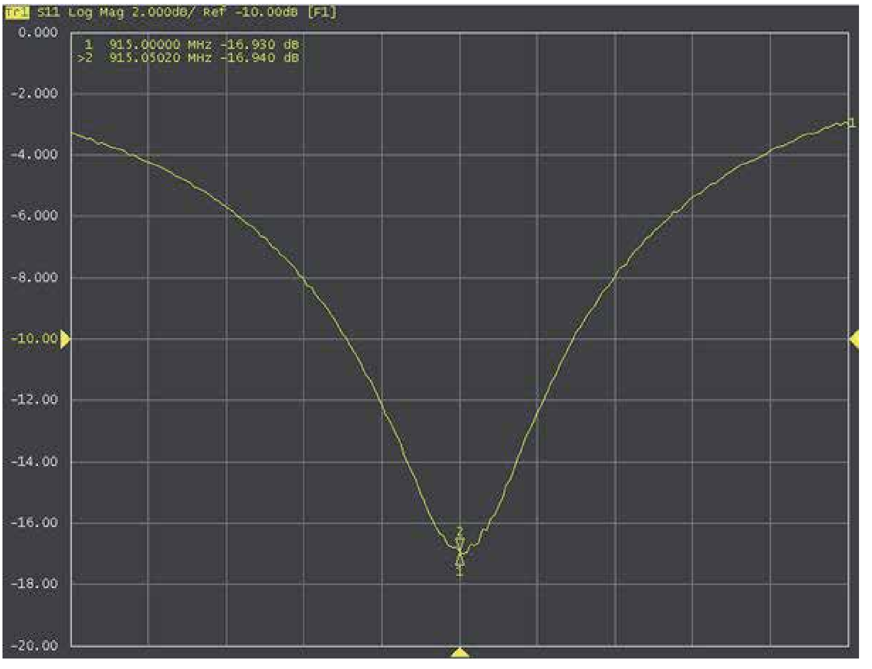
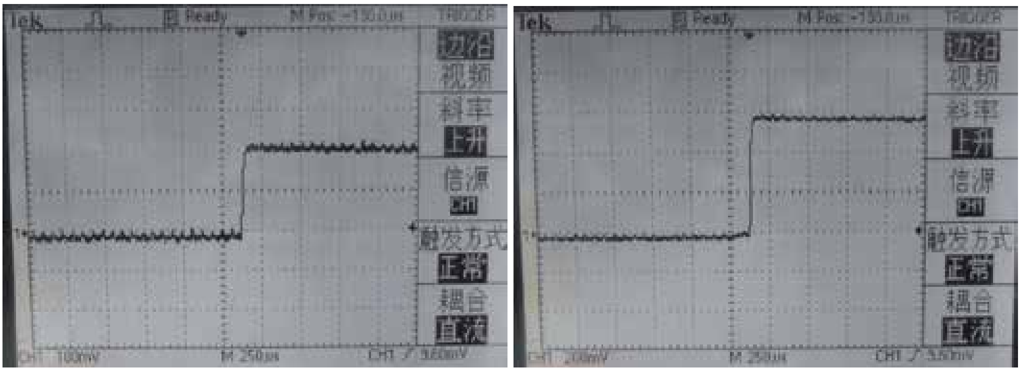


Figure 13. Measured return loss



(a) Input power of -27.7dBm (b) Input power of -21.7dBm

Figure 14. Measured DC output voltage with different input power

We also measure and simulate the output voltage of the circuit with different distance between the network controller and the WSNs node, and plot the results in the Fig. 16. We can see that the measured results are near to the simulation results too.

The above experimental confirm the performance of the proposed passive radio-triggered wake-up circuit, and the results demonstrate that the circuit can transform the RF signal to DC voltage efficiently. Compared with the schemes proposed in the literatures [3, 5, 9, 11, 14, 17], our proposed scheme is the first one that uses the discrete components to realize the design and achieves as good as or even better performance. It's cheaper and more flexible and can be applied to different WSNs applications with little change.

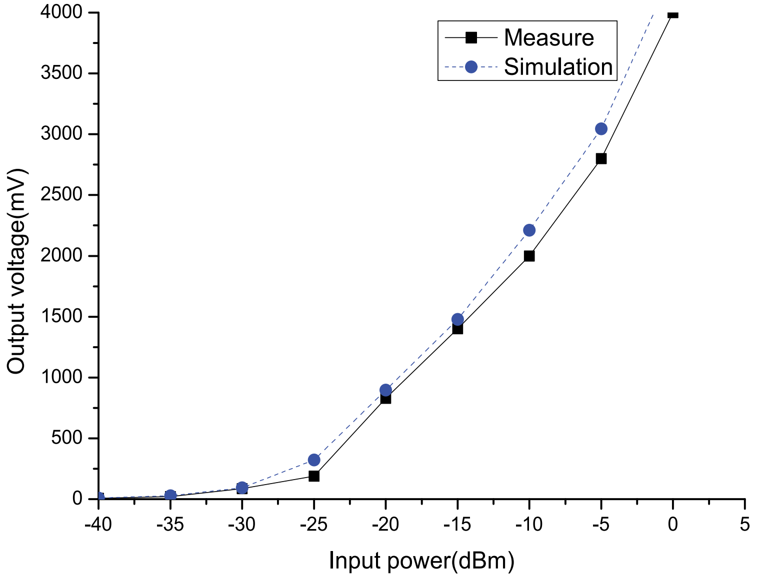


Figure 15. The measured and simulated output voltage as a function of the input power

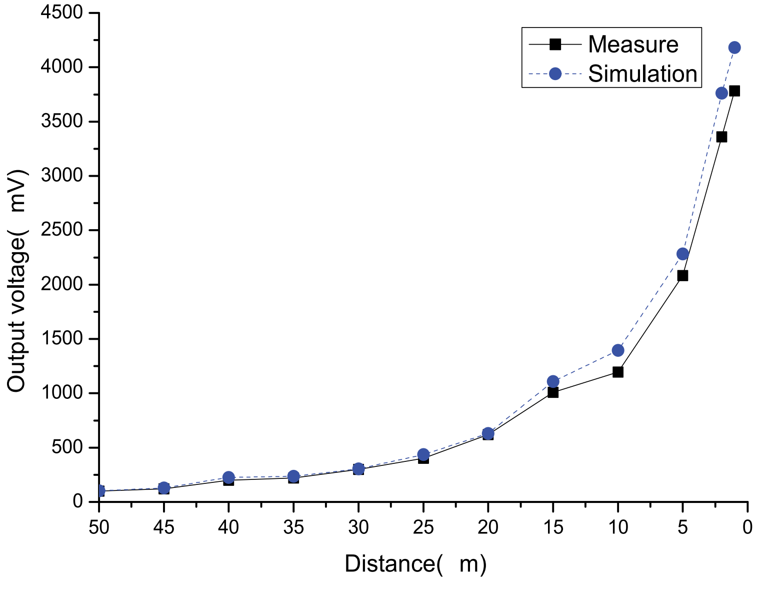


Figure 16. The measured and simulated output voltage as a function of distance

6. Discussion about some performance improvement schemes

As has been analyzed above, the proposed passive radio-triggered wake-up circuit could achieve reasonable performance and could be adopted by the WSNs node to realize the power management task. However, in some WSNs based applications, one may want to improve the wake up distance of the network controller, or want to realize the addressable wake-up so as to wake up one node while keep the other nodes in the sleep node. In this section, we present some performance improvement schemes for solving the above two problems. It should be mentioned that the improvement schemes may need extra power supply.

6.1. Long distance wake-up circuit

As has been discussed above, the two stages rectifier can produce a DC voltage of about 220mV with the distance of 40m. This is enough for triggering a micro-power low voltage supervisor. However, in some WSNs based applications, one may need a distance of 100m or more, and the rectifier can not produce a voltage high enough for meeting this requirement. In the literature, two advanced methods were introduced to tackle this problem [6]. One method uses a store-energy radio-triggered circuit where a transformer is used. The disadvantage of this method is that it needs a relatively longer time to realize wake-up since the transformer needs some time to store enough energy. Meanwhile, the size of the transformer is too large, which makes it improper to be used in the WSNs node. Another method uses an amplifier to amplify the DC output signal of the passive radio-triggered wake-up circuit. The amplifier has internal power supply, so it can generate an output signal higher than the input wake-up signal. The disadvantage of this method is that the amplifier still consumes energy when there is no wake-up signal. Typically, an amplifier draws about 1uA, and this is a little high for the WSNs application.

In this chapter, we propose a comparator based long distance wake-up circuit. As we know, a comparator typically draws less current than an amplifier. For example, Linear Technology's LT1540 ultra-low power comparator typically draws 0.3uA current. By adding a comparator to the output of the passive radio-triggered wake-up circuit, the DC voltage can compare with a predefined threshold voltage and produce a high voltage trigger voltage to interrupt the MCU. Compared with the amplifier based approach, the comparator based scheme consumes less energy. Hence, it is a good choice to design long distance wake-up circuit based on a comparator. The block diagram of the long distance radio-triggered wake-up circuit with a comparator is shown in Fig. 17.

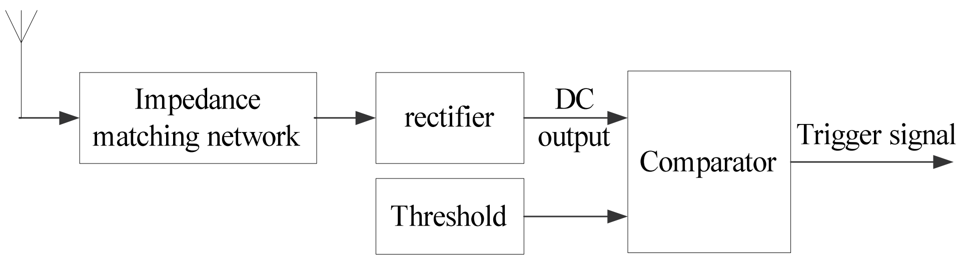


Figure 17. Long distance radio-triggered wake-up circuit with a comparator

In the Fig. 17, the threshold value setting circuit is used to change the working distance of the circuit. A low threshold value can be used to achieve a long working distance. The comparator compares the DC output of the rectifier with the threshold value, and produces a trigger signal to wake up the MCU. The circuit proposed in Fig. 17 can produce a trigger signal with a DC output from the rectifier as low as 10mV. The drawback of this circuit is the extra energy consumed by the comparator and threshold value setting circuit which is about 0.5uA in all. For some applications where very long working distance is required, this method is a good choice.

6.2. Addressable wake-up circuit

Essentially, the radio-triggered wake-up circuit is a simple receiver, one can utilize the multiple address technique that adopted in the traditional communication field to realize the addressable wake-up.

The most intuitive method is the frequency division multiple address scheme [4, 6]. A set of radio-triggered circuits which work on different frequencies are put on one node, and the network controller sends the wake-up signal on multiple frequencies at the same time. The outputs of the radio-triggered circuits are fed to the AND-gates so that the circuit could only wake up the node when all the corresponding frequencies are present. The shortcoming of this method is that the number of possible addresses is very limited. When using 6 different frequencies, only 64 different nodes can be distinguished. Another disadvantage of this method is that the cost increases significantly with the increase of possible addresses.

In this chapter, we propose a pulse width modulation based addressable wake-up circuit. Compared with the frequency division multiple address scheme, the pulse width modulation scheme may be an economic, simple, and efficient one. With the pulse width modulation scheme, only one frequency channel is required. When the MCU is waked up by the wake-up circuit, it measures the pulse width of the signals, and demodulates the address transmitted by the network controller. The principle of the pulse width modulation scheme is shown in the Fig. 18. The signals ’1’ and ’0’ have different pulse width, and the MCU can demodulate the signal simply with the help of an on-chip timer. In the Fig. 18, the waveform of the address of ’1010’ is also illustrated.

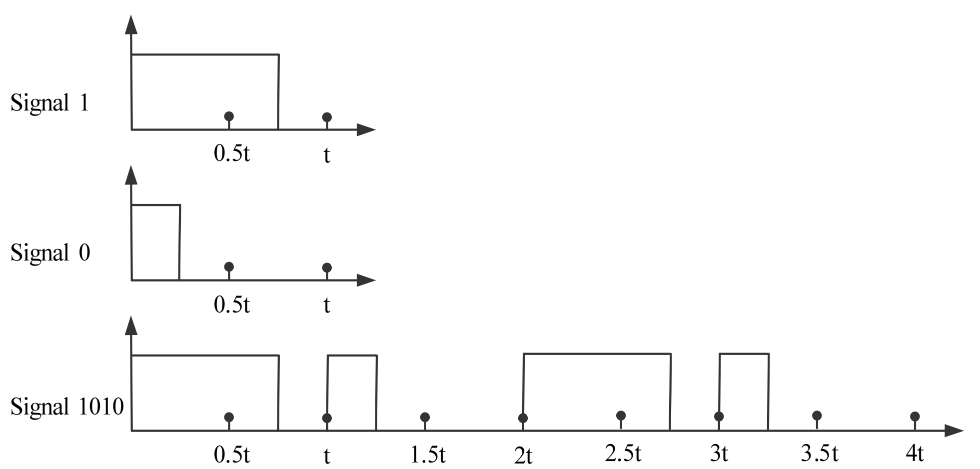


Figure 18. Principle of the pulse width modulation scheme

The advantage of this scheme is that it is very simple to be deployed in the WSNs node and network controller. There is no need to add any hardware, and the software for realizing the above scheme is so simple. Hence, the pulse width modulation based addressable scheme is an ideal choice for the radio-triggered wake-up circuit.

7. Conclusion and future work

In this chapter, we presented a radio-triggered wake-up circuit and explored its application in the power management of the WSNs. By harvesting energy from the radio signals, the radio-triggered hardware could provide a wake-up signal to the MCU without using power supply, and it takes no more than 30us for the circuit to produce the wake-up signal. The circuit could produce a DC output voltage of 220mV with the received power as low as 1.7uW (-27.7dBm), corresponding to a 40m distance in free-space with 4W radiation source. Meanwhile, we discussed some advanced schemes to improve the wake up distance and to realize the addressable wake-up. Equip with the radio-triggered wake-up circuit, the lifetime of the WSNs node could be prolonged.

It should be mentioned that the radio-triggered wake-up circuit can be used in other applications, such as the synchronization of the WSNs. And we will apply our radio-triggered wake-up circuit to the low duty cycle WSNs to realize the synchronization of the network. Meanwhile, we will explore other possible schemes for transforming the RF signal to DC voltage.

Acknowledgment

This work was supported by National Natural Science Foundation of China under grant numbers 61172058 and the National Hi_Tech Research and Development 863 program of China under grant numbers 2008AA092701. The authors would also like to thank the reviewers for their useful suggestions and comments.

Author details

Jie Wang, Qinghua Gao, Yan Yu, Hongyu Wang and Minglu Jin
Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, China

8. References

[1] Ameen, M. A.; Ullah, N. & Kyungsup, K. (2011). Design and Analysis of a MAC Protocol for Wireless Body Area Network Using Wakeup Radio, *Proceeding of the 11th International Symposium on Communications and Information Technologies*, pp. 148-153, ISBN 978-1-4577-1294-4, Hangzhou, China, Oct. 12-14, 2011

[2] Bahl, I. J. (2000). High-Q and Low-loss Matching Network Elements for RF and Microwave Circuits. *IEEE Microwave Magazine*, Vol. 1, No. 3, (Sep. 2000), pp. 64-73, ISSN 1527-3342

[3] Curty, J. P.; Joehl, N.; Dehollain, C. & Declercq, M. J. (2005). Remotely Powered Addressable UHF RFID Integrated System. *IEEE Journal of Solid-State Circuit*, Vol. 40, No. 11, (Nov. 2005), pp. 2193-2202, ISSN 0018-9200

[4] Cochran, T. L.; Kim, J. & Dong S. (2011). Low Power Wake-up Receiver with Unique Node Addressing, *Proceeding of the IEEE 54th International Midwest Symposium on Circuits and Systems*, pp.1 -4, ISBN 978-1-61284-857-0, Seoul, Korea, Aug. 7-10, 2011

[5] Janek, A.; Steger, C.; Preishuber-Pfluegl, J. & Pistauer, M. (2007). Power Management Strategies for Battery-driven Higher Class UHF RFID Tags Supported by Energy Harvesting Devices, *Proceedings of IEEE Workshop on Automatic Identification Advanced Technology*, pp. 122-127, ISBN 1-4244-1300-1, Alghero, Italy, Jun. 7-8, 2007

[6] Lin, G. & Stankovic, J. A. (2004). Radio-triggered Wake-up Capability for Sensor Networks, *Proceeding of the IEEE Symposium on Real-time and Embedded Technology and Applications*, pp. 27-36, ISBN 1080-1812, Toronto, Canada, May. 25-28, 2004

[7] Liang, S.; Tang, Y. & Zhu, Q. (2007). Passive Wake-up Scheme for Wireless Sensor Networks, *Proceedings of the Second International Conference on Innovative Computing*, Information and Control, pp. 507-507, ISBN 0-7695-2882-1, Kumamoto, Japan, sept. 5-7, 2007

[8] Miller, M. & Vaidya, N. (2005). A MAC Protocol to Reduce Sensor Network Energy Consumption Using a Wakeup Radio. *IEEE Transactions on Mobile Computing*, Vol. 4, No. 3, (May. 2005), pp.228-242, ISSN 1536-1233

[9] Philippe, L. & Roy, S. (2010). Low-Power Wake-Up Radio for Wireless Sensor Networks. *Mobile Networks and Applications*, Vol. 15, No. 2, (Feb. 2010), pp. 226-236, ISSN 1383-469X

[10] Rabaey, J. M.; Ammer, J.; Karalar, T.; Li, S.; Otis, B.; Sheets, M. & Tuan, T. (2002). Picoradios for Wireless Sensor Networks: The Next Challenge in Ultra-low-power Design, *Proceeding of the IEEE International Solid-State Circuits Conference*, pp. 200-201, ISBN 0-7803-7335-9, San Francisco, CA, USA, Feb. 7-7, 2002

[11] Shameli, A.; Safarian, A.; Rofougaran, A.; Rofougaran, M. & De Flaviis, F. (2007). Power Harvester Design for Passive UHF RFID Tag Using a Voltage Boosting Technique. *IEEE Tansactions on Microwave and Technoques*, Vol. 55, No. 6, (Jun. 2007), pp. 1089-1097, ISSN 0018-9480

[12] Saxena, S.; Mishra, S.; Sharma, A. K. & Chauhan, D. S. (2010). Review on Protocol based Approaches to Extend Lifetime of Wireless Sensor Networks. *International Journal of Computer Applications*, Vol. 11, No. 6, (Jun. 2010), pp. 23-29, ISSN 0975-8887

[13] Sthapit, P. & Pyun, J. (2011). Effects of Radio Triggered Sensor MAC Protocol over Wireless Sensor Network, *Proceeding of the 11th IEEE International Conference on Computer and Information Technology*, pp.546-551, ISBN 978-1-4577-0383-6, Pafos, Cyprus, Aug. 31-Sept. 2, 2011

[14] Umeda, T.; Yoshida, H.; Sekine, S.; Fujita, Y.; Suzuki, T. & Otaka, S. (2006) . A 950-MHz Rectifier Circuit for Sensor Network Tags with 10-m Distance. *IEEE Jornal of Solid-State Circuit*, Vol. 41, No. 1, (Jan. 2006), pp. 35-41, ISSN 0018-9200

[15] Wu, H. & Pan, Y. (2008). *Medium Access Control in Wireless Networks, Volume II: Practice and Standards*, Nova Science Publishers, pp.535-560, ISBN 978-1-60741-936-5, NY, USA

[16] Wang, J.; Gao, Q.; Wang, H. & Sun, W. (2009). A Method to Prolong the Lifetime of Wireless Sensor Network, *Proceeding of the 5th International Conference on Wireless Communications, Networking and Mobile Computing*, pp.1-4, ISBN 978-1-4244-3692-7, Beijing, China, Sep. 24-26, 2009

[17] Yi, J.; Ki, W. & Tsui, C. (2007). Analysis and Design Strategy of UHF Micro-power CMOS Rectifiers for Microsensor and RFID Application. *IEEE Transactions on Circuits and Systems*, Vol. 54, No. 1, (Jan. 2007), pp. 153-166, ISSN 1549-8328

Collision Free Communication for Energy Saving in Wireless Sensor Networks

Besem Abid, Hamida Seba and Souleymane M'bengue

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/46569

1. Introduction

A Wireless Sensor Network (WSN) distinguishes from other wireless or wired networks through its capability of interaction with the environment. Such networks have been proposed for various applications including search and rescue, disaster relief, smart environments, and localization systems. These applications require a large amount of battery-powered wireless sensors, and are generally designed for long-term deployments with no human intervention. Consequently, energy efficiency is one of the main design objectives for these sensor networks. The main causes of the energy wastage are:

- **Collision**: occurs when two or more nodes attempt to transmit a packet across the network at the same time. The transmitted packets must be discarded and then retransmitted, thus the retransmission of those packets increases the energy consumption and the latency.
- **Overhearing**: occurs when a node receives a packet destined to other nodes. Overhearing can be a major reason of energy waste mainly with a high node density causing a heavy traffic load.
- **Packet Overhead**: sending and receiving control packets consumes energy too and less useful data packets can be transmitted, since control messages does not contain any useful application data.
- **Idle listening**: it occurs when a sensor node listens to an idle channel to receive possible traffic. Usually a node in a WSN doesn't know when to wake up to receive a packet, thus it must keep its radio ON which consumes most of the energy.

Therefore, researchers give a growing interest on optimizing WSN MAC (Medium Access Control) to reduce the energy consumption of the sensors in order to extend the network lifetime. The main challenge of any MAC protocol is to avoid collision as it represents the most important issue of energy saving. Usually in WSNs several nodes share the same channel, thereby the probability of packet collision increases. Developing a MAC protocol to coordinate the channel access of these nodes decreases the risk of packet collision and specially DATA packet collision which decreases the channel utilization as DATA packets are longer than control packets.

In this paper, we provide a state-of-the-art study of WSN MAC protocols, and we will discuss the advantages as well as the drawbacks of the main existing solutions. We classify the MAC protocols according to the technique being used and to the problems they try to solve. In contrast to previous surveys, we will give more interest to the solutions treating the mobility of the sensor nodes and the real time constraints. Generally, MAC protocols are classified into two categories: contention based and schedule based protocols.

Contention based protocols allow many users to use the same radio channel without pre-coordination. The main idea of these protocols is to listen the channel before sending the packet, IEEE 802.11, ALOHA and CSMA (Carrier Sense Multiple Access) are the most well known contention-based protocols. Compared to the schedule based protocols, the contention one are simple, because they don't require global synchronization, or topology knowledge which allows some nodes to join or to left the network few years after deployment. Message collisions, overhearing and idle listening are the main drawbacks of this approach. Thereafter we will present these basic protocols (IEEE 802.11, ALOHA and CSMA), and in the next sections other contention based approaches more suitable for WSNs will be discussed.

- **IEEE 802.11**: to avoid collisions of data packets, IEEE 802.11 [1] uses carrier sensing and randomized back-offs. Likewise, the Power Save Mode (PSM) of the IEEE 802.11 MAC protocol is used to reduce the idle listening. PSM power management decisions are made periodically to allow the sensor nodes to switch to the sleep state. IEEE 802.11 PSM may not be suitable for multi-hop networks because of the problems of clock synchronization, neighbour discovery and network partitioning.
- **CSMA**: a node verifies the absence of other traffic on the shared transmission medium, and then if the channel is clear it starts transmitting the packet, otherwise the node waits for the transmission in progress to finish. CSMA is a robust method to nodes mobility and no central node is required, but it suffers from the hidden terminal problem which occurs when two transmitters are not able to sense each other because they are out of range of each other, so their messages might collide at the receiver node which is in range of both transmitters.
- **ALOHA**: the first version of the protocol was quite simple: if you have data to send, send it, then if the message collides with another transmission, try re-sending "later". The main drawbacks of this protocol is its poor use of the channel capacity; the maximum throughput of the ALOHA protocol is only 18% [2]. The channel utilization can be increased with slotted ALOHA (35%), by dividing time into slots and nodes may only start transmitting at the beginning of a slot. This method decreases the probability of packet collision.

  These protocols (ALOHA and slotted ALOHA) are not a suitable solution for multi-hop networks, and they do not solve the problem of hidden nodes which increases the problem of collision, and hence increases energy consumption.

In schedule based protocols, a schedule is established to allow each node to access the channel and communicate with other nodes. The main goals of this schedule based approach is to reduce collision and ensure fairness among the different nodes of the network, except that the network topology must be known in advance, and synchronization is required between nodes, thus the better the synchronization the shorter the idle listening. **TDMA (Time Division Multiple Access)** is a representative example for such a schedule based approach. It allows nodes to share the same frequency channel by dividing time into frames and each frame is divided into slots. The nodes transmit in rapid succession, one after the other, each using its own time slot; consequently transmissions do not suffer from collision. TDMA guarantees fairness among nodes as each node is assigned a unique slot in each frame, and increases the overall throughput in highly loaded networks.

In the next sections, we present the different contention based protocols suitable for WSNs, then we discuss schedule based protocols. We keep the last two sections to the protocols dealing with the mobility of the nodes and the applications with real time constraints.

2. Contention based MAC protocols for WSNs

As mentioned before contention based protocols require no coordination among the nodes accessing the channel, these protocols are mainly based on CSMA or CSMA/CA. To avoid the hidden terminal problem discussed above, RTS (Request To Send)/CTS (Clear To Send)/ACK (Acknowledgement) control messages are used. In general these protocols allow nodes to enter periodically into sleep periods. In the following, we describe the main contention based protocols proposed for WSNs and Table 1 summarizes them.

1. **S-MAC [3]**: is specially designed for wireless sensor networks and is derived from IEEE 802.11. A time frame in S-MAC is divided into two parts: one for a listening session and the other for a sleeping session, thus a sensor has a periodic listen/sleep period (S-MAC assumes that the duration of the listen and sleep periods is known by all nodes). Therefore, the nodes turn off their power-consuming transceiver and application packets are backlogged during their sleep period while the communication with other nodes is allowed during the listening period. A SYNC control packet is exchanged between nodes

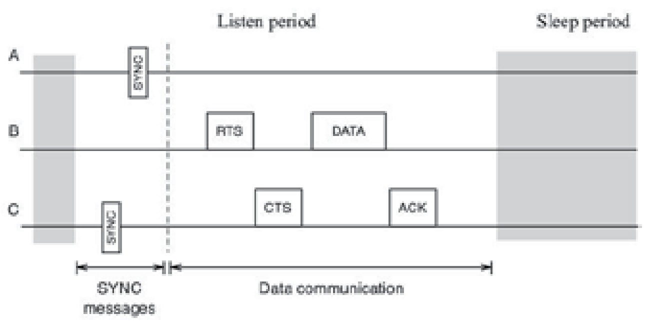


Figure 1. S-MAC messaging scenario [3]

during the listening period to establish synchronization among the neighboring nodes, i.e. schedule exchanges are accomplished by periodic SYNC packet broadcasts to immediate neighbors. SYNC messages are sent during synchronization period as shown in Figure 1.

S-MAC divides the listening period into two parts (Figure1: (1) the first part is reserved to the exchange of SYNC messages and (2) the second part allows the exchange handshaking mechanism (RTS/CTS) and DATA packets. To avoid packet collisions, S-MAC uses RTS/CTS mechanism similarly to IEEE 802.11. Moreover, it performs virtual and physical carrier sense before transmission. S-MAC tries to avoid overhearing by letting interfering nodes go to sleep after they hear an RTS or CTS packet. A node having a NAV value not equal to zero should go to sleep to avoid overhearing (note that NAV is the network allocation vector used to indicate the activity of every neighbor of a given node. When a node receives a packet destined to other nodes, it updates its NAV by the duration field in the packet).

The main advantages of S-MAC are:

- the energy waste caused by idle listening is reduced by sleep schedules.
- simplicity of implementation.
- time synchronization overhead may be prevented by sleep schedule announcements.

Apart from the above advantages, S-MAC suffers from some drawbacks such as:

- Sleep and listen periods are predefined and constant which decreases the efficiency of the algorithm under variable traffic load.
- Broadcast data packets do not use RTS/CTS, which increases collision probability.
- high latency and low throughput.

2. **T-MAC [5]**: is based on S-MAC [3], the main difference between the two protocols is that the active period in T-MAC is preempted if no activation event has occurred for a time TA, thus a sensor node could spend more time on sleeping mode compared to S-MAC. Figure 2 shows the overall operation of T-MAC and its difference compared to S-MAC. The interval $TA > C + R + T$ where C is the length of the contention interval, R is the length of an RTS packet and T is the turn-around time (the short time between the end of the RTS packet and the beginning of the CTS packet). The major problem of T-MAC [5] is

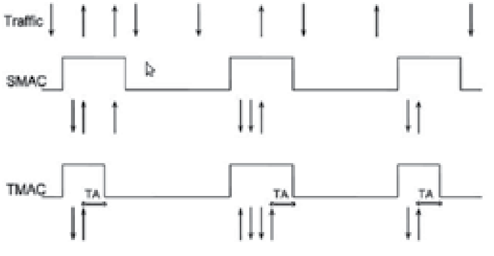


Figure 2. T-MAC and S-MAC active period operation [5]

the early sleeping problem, which occurs when a node goes to sleep while a neighbor still has messages. To overcome this problem the protocol uses Future Request-To-Send (FRTS) as a solution. An FRTS is sent to the neighbor node before its TA timer expires to force it to stay in active mode, thus it can receive the transmission at the same active period rather than receiving it in the next active period. The FRTS solution is illustrated in Figure 3.

3. **E2MAC [6]**: unlike T-MAC, it adds packets in a node's buffer until it reaches a maximum predefined size, then the node transmits data in a burst. This protocol increases energy saving, but it also increases latency and is not suitable for delay-critical applications.

4. **SWMAC (Separate Wakeup MAC) [7]**: in this protocol, the active period is divided into slots to reduce its duration, and then each node is assigned a reception slot during which

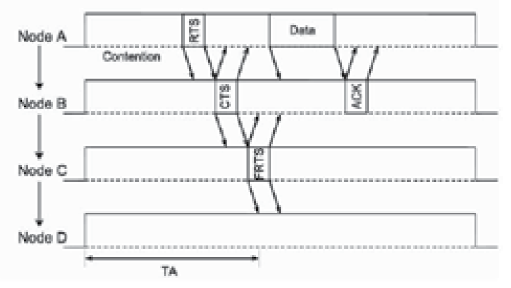


Figure 3. Future request-to-send scenario [5]

it wakes up to receive data. Apart from that, the protocol follows the same rules as S-MAC [3], thus SWMAC suffers from the same drawbacks as S-MAC.

5. **Adaptive Listening [8]**: it aims to minimize the *sleep delays* by the use of overhearing. In this protocol the node that hears another transmission could go immediately to the sleep state once it is informed about the transmission end of its neighbor.

   Note that the sleep delay means: end-to-end delay that increases latency in multi-hop networks, as intermediate nodes on a route do not necessarily share a common schedule.

6. **B-MAC [9]**: is highly configurable and can be implemented with a small code and memory size. To save energy, radios on B-MAC periodically wake up to sample the medium. B-MAC consists of four main parts: channel assessment (CCA) and packet backoffs for channel arbitration, link layer acknowledgments for reliability, and low power listening (LPL) for low power communication.

   B-MAC is based on preamble sampling like ALOHA protocol [2], the preamble length is configurable which gives optimal trade-off between energy savings and latency or throughput.

   The experimental results of [9] show that B-MAC has better performance in terms of latency, throughput and often energy consumption compared to S-MAC [3].

   With B-MAC, all nodes are in sleep mode when there is no traffic which represent the main advantage of this protocol, thus they just wake up only for preamble sampling and go back to sleep. However B-MAC is an efficient protocol only in low network traffic conditions since nodes will spend most of the time sleeping, otherwise, when traffic condition increases, B-MAC loses efficiency because nodes remain awake for a longer time waiting for the entire packet transmission.

7. **DSMAC (Dynamic S-MAC) [10]**: as the previous protocol, it aims to reduce the sleep delay by dynamically changing its duty-cycle depending on the application demands. When an application requires less latency or when there is an increasing traffic load, a node increases its duty cycle by adding extra active periods. In this case, the concerned node sends a SYNC packet to its neighbors to inform them about its additional active schedule, then each neighbor decides to increase its duty-cycle or not. At any time, the sensor node can decrease its duty cycle by removing the added active periods.

8. **U-MAC [11]**: is based on S-MAC [3] and provides three main improvements on this protocol: various duty-cycles, utilization based tuning of duty-cycles and selective sleeping after transmission. Each node of the network has a different periodically listen

and sleep schedules with different duty cycles. Like S-MAC, nodes have to exchange their schedules and be synchronized. However they do not adopt the same schedules as their neighbors. Additionally, the ACK packet contains always the time of the next sleep of a node. Utilization based tuning of duty-cycles reflects different traffic loads of every node in a network. Such variation corresponds to the diversity of performed tasks by a particular node and its location. Selective sleeping after transmission avoids the above energy wastage. After the transmission of a packet, a node goes to the sleep state if it finds a scheduled sleep time. It does not introduce additional delays, since no traffic is expected to this node.

U-MAC [11] improves energy efficiency as well as end-to-end latency compared to the previous protocols. As shown in Figure 4 and Figure 5, U-MAC saves about 32% energy and 45% latency from S-MAC.

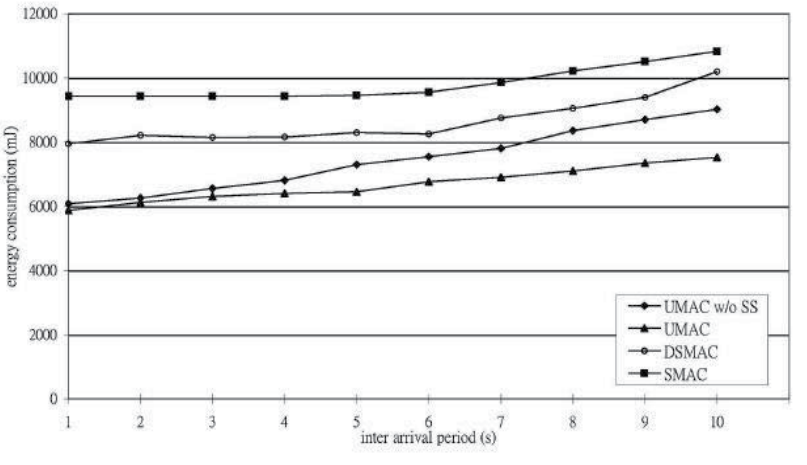


Figure 4. The energy consumption of U-MAC without sleep after transmission (U-MAC w/o SS), U-MAC with selective sleeping after transmission (UMAC), DSMAC, and S-MAC [11]

Contention based protocols are known also as random access protocols. The main drawbacks of such protocols are the performance degradation when traffic is frequent and these protocols suffer from stability problems.

3. Schedule based MAC protocols for WSNs

These protocols are based on a schedule to carry out communication. They are also called TDMA based protocols. Using this approach, we achieve a good energy efficiency and throughput [12]. Although, schedule based protocols require synchronization between nodes to guarantee a shorter idle listening.

In the following we will present the main schedule based protocols proposed for WSNs and we summarize these protocols in Table 2.

1. **MMF-TDMA [13]**: the main goal of Max-Min Fair TDMA (MMF-TDMA) approach is to assign bandwidth to nodes using linear programming min-max fairness approach.

Protocol Name	Scheme Used	Energy Saving	Advantages	Disadvantages
S-MAC [3]	fixed duty cycle, virtual cluster, CSMA	power savings over standard CSMA/CA MAC	low energy consumption when traffic is low	sleep latency, problem with broadcast packets
T-MAC [5]	adaptive duty cycle, overhearing, FRTS	uses 20% of energy used in S-MAC.	adaptive active time	early sleeping problem
U-MAC [11]	various duty-cycles, utilization based tuning of duty-cycle, selective sleeping after transmission	performs better than S-MAC	improves energy efficiency and end-to-end latency	early sleeping and clock drift problems
E2MAC [6]	same as T-MAC and it accumulates packets in a node until they reach a certain buffer size limit	performs better than S-MAC	increases energy savings	increases latency
Adaptive Listening [8]	suggests the use of overhearing to reduce the sleep delay	performs better than protocols without adaptive listening	minimizes the sleep delays	problem with broadcast packets
DSMAC [10]	adaptive duty cycle	performs better than S-MAC	reduces the sleep delay	can only decrease node-to-node delay and is not suitable to enforce timely data delivery. Problem of clock drift.
B-MAC [9]	based on preamble sampling	nodes are in sleep mode when there is no traffic	has better performance in terms of latency, throughput and often energy consumption as compared to S-MAC	when traffic conditions increase, B-MAC loses efficiency.

Table 1. Summary of contention based protocols in WSNs

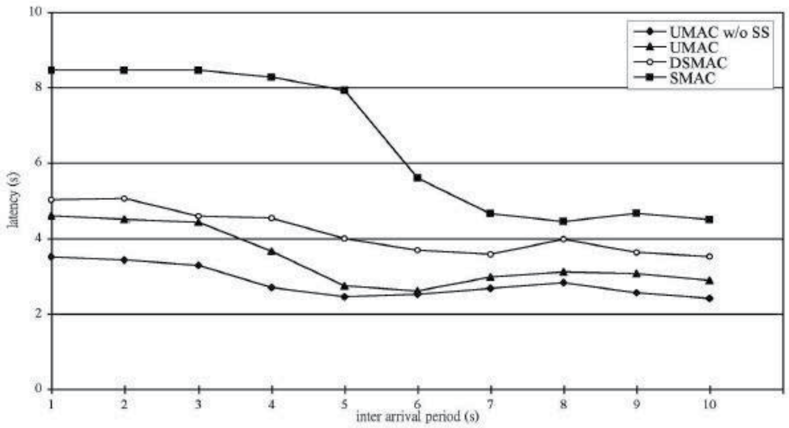


Figure 5. The latency of U-MAC without sleep after transmission (U-MAC w/o SS), U-MAC with selective sleeping after transmission (U-MAC), DSMAC, and S-MAC [11]

MMF-TDMA approach considers the network as a tree-like network organization where the parent nodes coordinate the assigning process of their siblings.

To guarantee collision-free, [13] propose a distributed max-min fair scheduling mechanism suited for continuous traffic on a data gathering tree and incorporate it with a time-slot based bandwidth allocation scheme. For max-min fair resource allocation, if the source nodes of the network are not constrained, they increase their allocation by a small value for every iteration of the algorithm. When the combination of incoming data, generated data, and interfering data traffic is within the total bandwidth, we said that those nodes are constrained. Note that all nodes having an output traffic interfere at a constrained node and nodes that are in the subtree below a constrained node, become constrained. The algorithm ends when all nodes on the routing tree become constrained and the rate available to all sources is the allocated rate.

A TDMA based scheduling algorithm is used to provide enough number of time slots for the data originating at each source for each frame. The time slot allocation algorithm must avoid the hidden terminal problem, by ensuring that a time slot allocated to a node does not collide with that allocated to a node that is 2 hops away and that might interfere at the recipient of this node's communication. First, the root node allocates the required number of time slots to each of its children. Then, the time slot allocation algorithm runs in an iterative manner in a BFS-order (breadth first search). At each iteration, only one node allocates time slots to its children.

Simulations show that MMF-TDMA outperforms an overhearing avoidance MAC (similar to S-MAC) and pruned 802.11 (without ACK's) in terms of energy consumption, fairness, and delay. Although, this approach is not suitable for mobile nodes.

2. **D-MAC**: the data-gathering MAC (D-MAC) [14] aims to achieve a very low latency and still be energy efficient, for converge cast communication. It is designed for tree based data gathering.

   This approach is based on TDMA where time is divided into small slots. Within each of these slots, to transmit or to receive packet, CSMA is executed with acknowledgement. D-MAC guarantees a delay of approximately tens of milliseconds: A packet sent from a source node at a depth k can reach the sink node with a delay of just k time slots as shown in Figure 6. With D-MAC, a node has 3 states: sending, receiving, and sleeping.

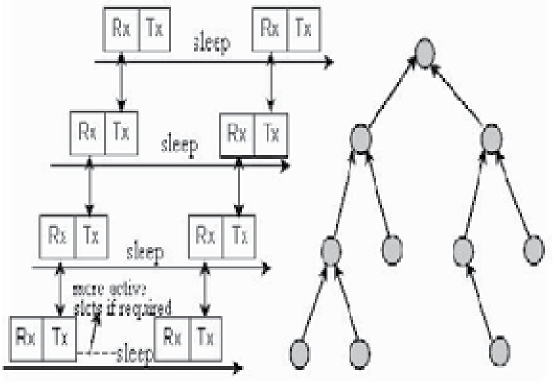


Figure 6. Data gathering tree in the D-MAC scheme [14]

During the first state a node sends a packet to the next hope on the route and awaits an acknowledgment. When the next hope node receives the packet (it is in the receiving state at the same time when the first node is in the sending state), it immediately switches to the sending state and sends the packet to the next hope. The sleep state of the node is the time interval between two successive sending and receiving periods as shown in Figure 6.

To reduce packet collision which occurs when two nodes in the same depth in the tree have synchronized schedules, D-MAC uses a back-off period plus a random time within a contention window.

The main problem of this approach is packet collision which occurs because many nodes in the tree could share the same schedule and D-MAC employs restricted collision avoidance methods. D-MAC works better for networks in which the transmission paths and rates are well known and do not change over time.

3. **TRAMA [15]**: the main goals of TRAMA (Traffic-Adaptive Medium Access) are the energy efficiency and network throughput. To achieve the first goal, the protocol aims to ensure collision free transmission and switching the nodes to idle state when they are not transmitting or receiving. There are three main parts in this protocol: the neighbor protocol (NP), the schedule exchange protocol (SEP) and the Adaptive election algorithm (AEP).
   - **Neighbor protocol**: is used to exchange one-hop neighbor information. This algorithm guarantees that the slots are long enough to allow all nodes to get consistent two-hop neighbor information.
   - **Schedule exchange protocol**: this part of the protocol is used to two-hop neighbor information and their schedule. Each node of the network computes a duration named $SCHEDULE - INTERVAL$ (the number of slots for which the node can announce its schedule to its neighbors). At time t the sensor node announces the slots where it has the highest priority among its two-hop neighbors. The selected slots must be in the following time interval $[t, t + SCHEDULE - INTERVAL]$.
   - **Adaptive election algorithm**: ensures that there is a unique ordering of node priorities within any two-hop region at each time, thus a node transmits its packet only when it has the highest priority among the two hop neighbors.

The main problem of this protocol is its long delay compared to the other protocols, which make it suitable only for applications that are not delay sensitive but require higher energy efficiency and throughput. Figures 8 (a) and 8 (b) show respectively the percentage of energy savings and the percentage of sleep time achieved by TRAMA and S-MAC in three networking scenarios illustrated in Figure 7. We can see that TRAMA has better performance than S-MAC especially in the edge initiators scenario [15].

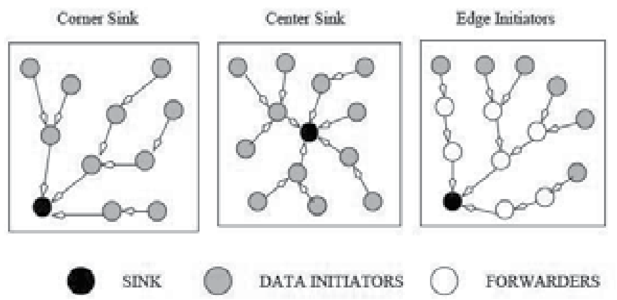


Figure 7. Data gathering application [15]

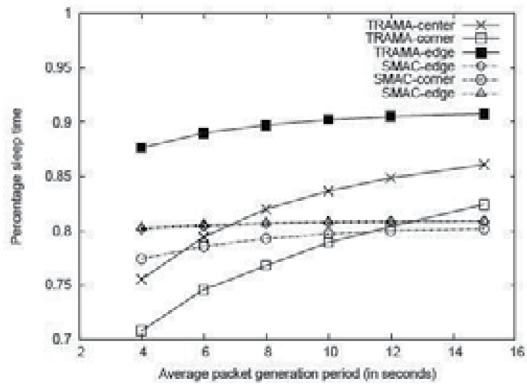
4. **Energy Aware TDMA Based MAC (EATDMA)** [16]: the network is divided into clusters, where each one of them is managed by a gateway. The main functions of the gateway are: collects information from the other sensor nodes within its cluster, performs data fusion, communicates with the other gateways and sends data to the control center, assigns time slots to the sensor nodes within its cluster and informs the other nodes about the time slot when they should listen to other nodes and the time slot when they can transmit own data. This approach has four phases:
   (a) **Data transfer**: sensors send and relay data in the allocated time slots.
   (b) **Refresh phase**: each node informs the gateway about its energy level, state, position, etc...
   (c) **Event-triggered rerouting**: a gateway runs the routing algorithm based on the information sent by the nodes during the previous phase.
   (d) **Refresh-based rerouting**: the gateway performs this action periodically after the refresh phase.

   With EATDMA[16], the node, that only senses, receives one time slot, while it receives as many time slots as nodes send information through it, if the node relays data. The Node that senses and relays receives one time slot for itself and as many time slots as the nodes that send information through it.

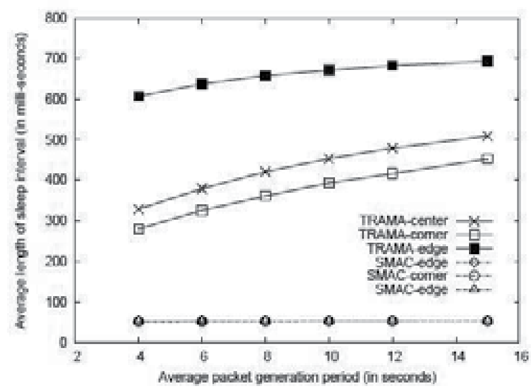
   There are two approaches for slot assignment: Breadth First Search (BFS) and Depth First Search (DFS). The former assigns the time slot numbers starting from outer most sensor nodes and giving them contiguous slots. While the latter assigns contiguous time slots for the nodes on the route from outermost sensor nodes to the gateway.

   Simulation shows that BFS technique gives better results with respect to DFS in terms of node lifetime. However, DFS has low latency and high throughput as compared to BFS.

5. **PEDAMACS [17]**: in Power Efficient and Delay Aware Medium Access Control protocol for Sensor networks of [17], the schedule is calculated and sent to the entire network by



(a) Percentage Energy Savings



(b) Average Sleep Interval

Figure 8. Energy savings and average sleep interval for sensor scenarios [15]

the sink node based on the information about the network topology and traffic load. That information are sent by the other nodes of the network.

At the end of the topology learning phase a spanning tree is constructed and the sink node has an entire knowledge of the topology. This phase is done by the topology-learning packets flooded in the network by the sink node.

The main drawback of this protocol is that the traffic pattern is always converge cast. In addition the protocol assumes that the sink is powerful enough so that it can reach all nodes when it transmits which is not always true, thus, nodes that do not receive the schedule transmitted by the sink, must wait for the next topology-learning.

6. **G-MAC [18]**: divides the message frame into two periods: the collection period, and the distribution period. During the first period, the gateway sensor node (the cluster

coordinator), collects information sent by the other nodes of the cluster expressing their future upload traffic. In the distribution period, the gateway sensor node sends a GTIM message (gateway traffic indication message) to the other nodes. GTIM maintains synchronization among nodes and sets up slot owners among nodes having data to be sent to the gateway. G-MAC periodically elects a new gateway node to equally distribute the energy requirements among all of the sensors. Compared to S-MAC by [3], and T-MAC by [5], G-MAC protocol provides the best network lifetime, but it still has a large overhead.

7. μ**MAC**: this protocol proposed by [19] uses a similar idea as TRAMA [15]. It has two periods: the contention period for two-hop topology construction, and a contention free period for data exchange. Other forms of communication take place in special reserved time slots for broadcasting (not contention-free) communication. Compared to TRAMA described by [15], μMAC uses a different time slot reservation mechanism.

Schedule-based protocols are energy efficient, as the schedule forces nodes to switch on only during a specific time slot, for the rest of the time they are in sleep mode. With respect to the contention based protocols, scheduled protocols solve the problem of interference and reduce packet collision. However, this family of protocols suffer from several drawbacks, as the limited scalability and flexibility due to the frequent topology change in WSN, and they perform worse than contention based protocols in low traffic condition. To overcome the drawbacks of the above protocols researchers propose hybrid solutions to combine the strengths of both scheduled and contention based protocols.

4. Hybrid solutions

These protocols take advantages of the above discussed protocols by exploiting the scalability and low control overhead of contention-based protocols and the high channel utilization efficiency of scheduled protocols.

Data communication occurs in both contention-based and scheduled fashion to achieve high performance under variable traffic load, thus, the former guarantees high performance when a small number of nodes transmit data in the network. However the latter achieves high performance when a large number of nodes transmit. The different protocols discussed in this section are summarized in Table 3.

1. **IEEE 802.15.4**: this standard [20] provides two services: the MAC data service to enable the transmission and reception of MAC protocol data units (MPDU) across the PHY data service and the MAC management service. The features of MAC sublayer are beacon management, channel access, GTS (Guarantee Time Slot) management, frame validation, acknowledged frame delivery and association and disassociation. The standard defines two types of network nodes: The full function device (FFD) that can serve as a network coordinator. This type of node has the possibility to talk with any other node, and to form any type of network topology. The second type of nodes is named reduced function devices (RFD). They are very simple devices and can form only the star topology when they are connected to a network coordinator (FFD). IEEE 802.15.4 is very flexible and defines two communication modes: The beacon-enabled mode and the non beacon mode. The first one is based on the super-frame shown in Figure 9, it contains an active and an inactive period. Although the active period contains 16 time slots and it is divided into a contention access period (CAP) and an optional contention-free period (CFP). The

Protocol Name	Scheme Used	Advantages	Disadvantages
MMF TDMA [13]	assigns bandwidth to nodes using linear programming min-max fairness approach	provides improvements in terms of key metrics such as energy, fairness, throughput, and delay	Is not suitable for mobile nodes in the network and it supports only tree-topology
D-MAC [14]	each node determines its active schedules based on the traffic load and its depth in the tree	D-MAC achieves very good latency	collision avoidance methods are not utilized
TRAMA [15]	determines a collision-free scheduling and performs link assignment according to the expected traffic	higher percentage of sleep time and less collision probability are achieved, as compared to CSMA-based protocols	the duty cycle is at least 12.5 percent, which is a considerably high value
EATDMA [16]	divides the network into clusters and time slots assignment is based on the breadth and depth techniques	the different techniques give better results in terms of node life time, latency and high throughput.	the main issue of this protocol is its scalability
PEDAMACS [17]	the schedule is calculated and sent to the entire network based on the information about network topology and traffic load	is power efficient and delay aware	the traffic pattern is always convergecast and the sink node cannot reach all the other nodes
G-MAC [18]	divides the message frame into two periods: the collection period, and the distribution period	compared to S-MAC [3], and T-MAC [5], G-MAC provides the best network lifetime	still has a large overhead

Table 2. Summary of schedule based protocols in WSNs

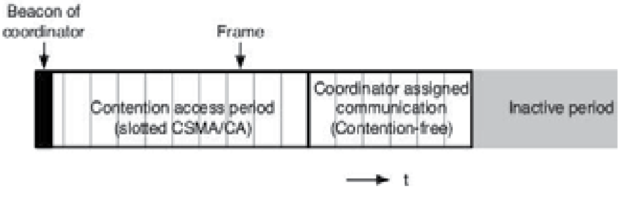


Figure 9. Superframe structure of the IEEE 802.15.4 MAC [20]

super-frame reserves a maximum of 7 time slots of the 16 time slots to the CFP TDMA communication.

The CSMA/CA protocol is used during the CAP period to access the channel and the beacon message of the coordinator specifies at which time slot the contention access period ends.

The CFP contains a special time slot assigned by the coordinator. Therefore, a sensor node can communicate with the coordinator. The standard specifies that the guaranteed time slots (GTS) can only be assigned by the PAN coordinator. In the following we will describe how data is transmitted from and to the coordinator.

- **Data transfer to a coordinator**: to transfer data to the coordinator a device can use the contention free period if it has a GTS, or it has to compete with other devices to access the channel during the contention access period. In the latter case, each device uses CSMA/CA procedure, then, if no collision occurrs, the coordinator sends an acknowledgement message to the device without using CSMA/CA protocol.

  The main problem of this approach is that the coordinator must be in listen mode during the contention period, to overcome this problem the IEEE 802.15.4 standard introduces a battery lifetime extension mode that limits the interval in which the coordinator listens to the channel after it has transmitted a beacon message. The standard specifies that in this energy-conserving mode, the coordinator stops listening to the channel after 6 back-off periods. However, it will still serves the GTSs as specified in the beacon message.
- **Data transfer from a coordinator**: in this case, using a beacon message, the coordinator informs the end devices that it has data pending. The concerned sensor node has to request for the data by sending a request message. If no collision occurs, the coordinator sends an acknowledgement and the data are transmitted using the slotted CSMA/CA mechanism. The end-device remains in receive mode until the data is received.

2. **Zebra MAC** (Z-MAC) [21]: uses CSMA in low traffic and switches to TDMA in high traffic conditions. Once the network is deployed, Z-MAC starts with a discovery phase of two-hop neighborhood followed by a slot assignment to nodes using DRAND (Distributed Randomized TDMA Scheduling For Wireless Adhoc Networks) by [22]. DRAND is a distributed protocol used to guarantee that a time slot is not assigned to two nodes located within three hops from each other. The algorithm takes care to distribute slots in a way that avoids hidden node collisions which may happen when a node and its two-hop neighborhood share the same time slot.

   To access the medium, if the node owns the current slot, it waits a random time smaller than a value T_o, called the owner contention window size, then performs a CCA (Clear Channel Assessment). If the channel is free, it emits. Otherwise, it waits until the channel becomes free again and resumes the same approach. If the current slot belongs to a two-hop neighbor and if the node has received an indication of strong contention from one of its two-hop neighbors, the node has no right to use this slot. Otherwise, it waits for a random time between T_o and T_{no} (the non owner contention window size) before performing a CCA.

   To evaluate the performance of Z-MAC, the authors of [21] compared Z-MAC and B-MAC [9] with $T_o = 8$ and $T_{no} = 32$. The default initial and congestion backoff window sizes of B-MAC are 32 and 16 slots respectively (each slot is 400 μs). The simulation results show that the power consumption of Z-MAC is slightly worse than that of B-MAC due to clock synchronization messages and the wake up period (see Figure 10).

3. **Funneling-MAC (FMAC)** [23]: takes into account the bottleneck faced by most of sensor network applications. This phenomenon occurs when a station in the network acts as a sink of data, to which a set of sensors direct their traffic. Funneling-MAC adopts an access method based on CSMA/CA in the entire network during a time interval followed by a time interval during which a TDMA access method is used for high-load zone only to

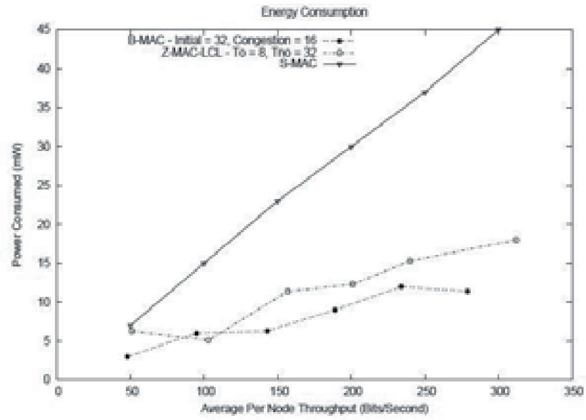


Figure 10. Power efficiency of Z-MAC in low-data rate applications with low duty cycle [21]

offer more access time to the nodes near the sink. These two time intervals constitute a superframe. This division is shown in Figure 11.

Area at high load is sized by a beacon message broadcast by the sink node. The nodes that do not receive the beacon apply the CSMA/CA procedure.

Based on the path of the received frames, the sink node determines the sequencing and the design of TDMA slots allocated to nodes in the area with a high load. Only the nodes belonging to the area with a high load update the path taken by a frame.

To avoid the interference between nodes beyond the high load area, the time slicing of TDMA, and the beacon transmitted by the sink node; nodes in the area with a high load periodically generate a frame containing information about the CSMA/CA period, the duration of the TDMA period, and the number of superframes until the next beacon.

Hybrid protocols try to combine the strengths of two protocols family (contention based and scheduled based). However these solutions are complex in terms of deployment.

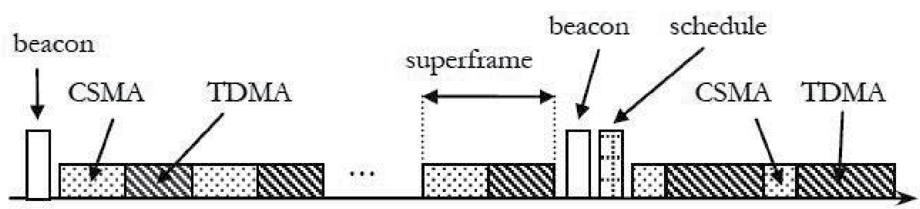


Figure 11. Funneling MAC: superframe [23]

5. MAC protocols for mobile sensors in WSNs

Most of the previous protocols deal with stationary sensors, whereas the new WSN applications use mobile nodes as they become essential for various areas such as: supply chain management, patient or children monitoring, and mobility platform in battlefield surveillance.

Protocol Name	Scheme Used	Advantages	Disadvantages
IEEE 802.15.4 [20]	provides two services: MAC data service and reception of MAC data units	very flexible by allowing to switch between many different modes	the number of frequency channels specified for IEEE 802.15.4 does not suffice to operate a variety of collocated WPAN applications that the standard is targeting
Zebra MAC [21]	uses CSMA in low traffic and switches to TDMA in high traffic conditions	increases the throughput	problem of schedule drift which reduces its energy efficiency.
Funneling MAC [23]	uses TDMA in regions close to the sink and CSMA elsewhere	very flexible	suffers from network dynamics

Table 3. Summary of hybrid protocols in WSNs

Any MAC protocol developed to handle node mobility must deal with topology changes Therefore, it should adapt neighbor table and route maintenance to the mobility in the network. Eventually, the MAC protocol needs mobility information, and must determines the neighborhood of each node to eliminate the inconsistency caused by the mobile node when they enter or leave the neighborhood.

In the following we introduce the main protocols that handle node mobility in WSNs and Table 4 summarizes them.

1. **SMACS and EAR algorithm [24]** SMACS(Self-organizing Medium Access Control for Sensor networks) is a distributed infrastructure-building protocol that allows sensor nodes to discover their neighbors and to establish the schedule of transmission and reception based on TDMA scheduling with no need of a master node. SMACS assumes that the network is connected, thus there exists at least one multihop path between any two distinct nodes. The Eavesdrop -And- Register (EAR) algorithm is based on SMACS. The EAR algorithm can realize the reliable communication between a mobile node and a fixed node. This communication is realized by the control head packets, which must be as low as possible. The main problem of this algorithm is related to the number of mobile nodes in the network; when this number is high it leads to packet collision with a high probability rate. It also does not guarantee high rate of coverage in the monitoring area. Note that the EAR algorithm is transparent to the SMACS protocol.

   In the EAR algorithm there are four types of frames to build a link between mobile nodes and stationary nodes: Broadcast Invite **(BI)**, Mobile Invite **(MI)**, Mobile Response **(MR)** and Mobile Disconnect **(MD)**. A BI frame is used by a stationary node to invite a mobile node to join a communication, thus a mobile node starts its connection protocol when it receives the BI frame. This frame is mainly used to register a stationary node depending on the connection status of the mobile node and the link quality between the mobile node and the stationary node. The mobile device will continue the registration procedure until its registry becomes full (any new stationary node will enter the register only if its link quality is better than the inferior registered link quality).

An MI frame is used as a response to a BI frame and a request to build up a connection. When the stationary node receives the MI frame, it decides whether the connection is possible or not. In the former case, slots are selected along the TDMA frame for communication, and a reply is sent to the mobile node accepting the connection.

the mobile node sends an MD frame to inform a stationary node to interrupt linkage. This case happens when the received SNR (signal to noise ratio) degrades and becomes lower than a given threshold. After a stationary node receives this frame, it deletes information of the mobile node from the registration form.

EAR algorithm can be used only in WSNs based on TDMA technology, which represent the main drawback of this algorithm. Furthermore, the network topology could only be cluster-based. Because of the important number of BI frames sent by the stationary nodes, energy efficiency of these nodes is low.

2. **Mobile sensor MAC (MS-MAC)** [25]: is based on S-MAC [3]; it varies the sleeping time dynamically according to the velocity of the mobile node. The first phase of this algorithm aims to build up virtual clusters by sending SYNC frame like in S-MAC. This phase of the algorithm is executed at the beginning at every scheduling mechanism.

   The main problem of this algorithm occurs when nodes move from one cluster to another, and it takes two minutes for a mobile node to get connected to a new cluster. To overcome this problem, border nodes should follow synchronous periods of the two virtual clusters, and set a value v_0 if the velocity of the mobile node is higher than $1/4\ v_0$, make scheduling cycle as 1 minute if the velocity of the mobile node is higher than $1/2\ v_0$, then make scheduling cycle as 30 seconds, etc. This mechanism of synchronization period is shown in Figure 12

   From the above algorithm discussion, we could conclude that MS-MAC cannot ensure reliable communication between stationary nodes and mobile nodes; meanwhile, this kind of sleeping mechanism also does not ensure a high rate of coverage and connection of the whole network.

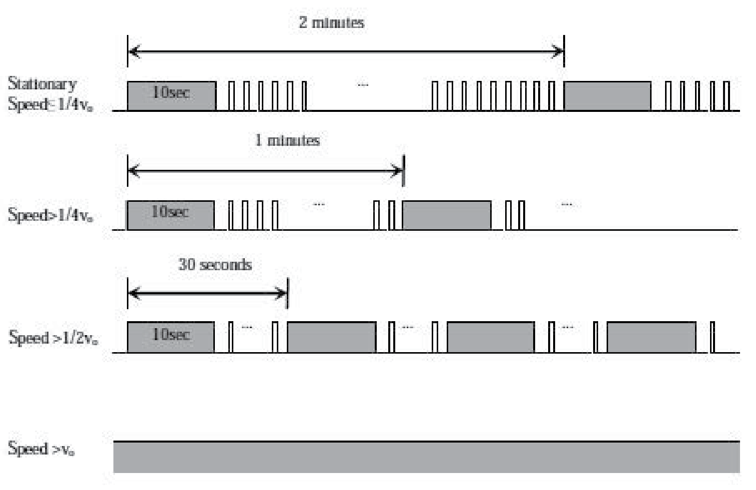


Figure 12. Frequency of synchronization period in MS-MAC depends on mobile speed [25]

3. **Mobility adaptive MAC (MMAC)** [26]: handles weak and strong mobility. The former means concurrent node joins and failures, and physical mobility either because of mobility in the medium (e.g. water or air) or by means of special motion hardware. While the latter means topology changes (node joins, and node failures).

   MMAC is dynamically adapted to changes in mobility patterns by introducing a mobility-adaptive frame time and the protocol builds a collision-free schedule based on estimates of traffic flow, mobility and dynamic patterns. The frame time is shown in Figure 13. To predict the mobility behaviour of sensor nodes and to adjust the frame time, MMAC uses the location information of the sensors. In Figure 13 we note that the frame time is dynamic and different from one node to another differently from the previously discussed TRAMA protocol of [15]. The basic idea of the Mobility-Adaptive algorithm is to reduce

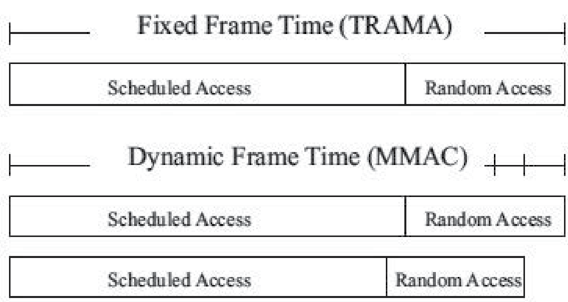


Figure 13. Fixed frame time (TRAMA) vs Mobility adaptive dynamic frame time (MMAC) [26]

   or to raise the frame time depending on the number of nodes that are expected to enter or leave the two-hop neighborhood of a given node. Moreover, the algorithm builds two sets: incoming nodes (IN) and outgoing nodes (ON) for a given node β, then if a node $'A'$ is a member of either $'IN'$ or $'ON'$ do not consider $'A'$ in the 2-hop neighborhood of β. To reduce the frame time the number of members in the previous two sets must be greater than a threshold value. Finally according to the frame time the algorithm adjusts the Scheduled-Access time and the Random-Access time shown in Figure 13.

   The main issues of this protocol are: mobility information (each node requires future mobility states of all current and potential two-hop neighbors) and Synchronization (each node has its own frame time which causes synchronization problems). To handle these problems MMAC proposes the following solutions:

   - **Mobility Information**: the signal and the data header are modified to include predicted mobility-state information, thus at the start of each frame, each node β independently calculates the expected mean (x,y) position of β in the next frame, then this information is sent in the header of every signal and data packet generated by β. The last scheduled access slot of the frame time is reserved for BROADCAST from head node for sending all received mobility information to the member nodes, therefore, each node β has 'best-effort' knowledge of the predicted mobility states of its current and potential two-hop neighbors.
   - **Synchronization problem**: to solve the synchronization problem, [26] introduces the concept of "Global Synchronization Period" (GSP) where the frame times would change only during this period. GSP occurs before each LEACH-style round (a mechanism to select cluster heads in MMAC) when cluster-heads are re-elected. During a round

(i.e. k frames), some changes may happens in the mobility rate, thus MMAC alters the division between scheduled access and random access slots after each frame. Therefore, the frame time in the network remains the same but the random access period of each cluster members would increase or decrease reflecting the mobility pattern. If all two-hop members of a node α belongs to a cluster c, then their random access time and scheduled access time would be the same.

The main disadvantage of this protocol is the requirement on the knowledge of the position, which is often either not feasible or too energy consuming.

Figure 14 shows that MMAC performs slightly better than TRAMA [15], while it outperforms S-MAC [3] and CSMA. In Figure 15, we can see that when the mobility of nodes is minimal all the protocols (apart from CSMA) are energy efficient. However, when the mobility becomes important only MMAC adapts to the mobility of the nodes.

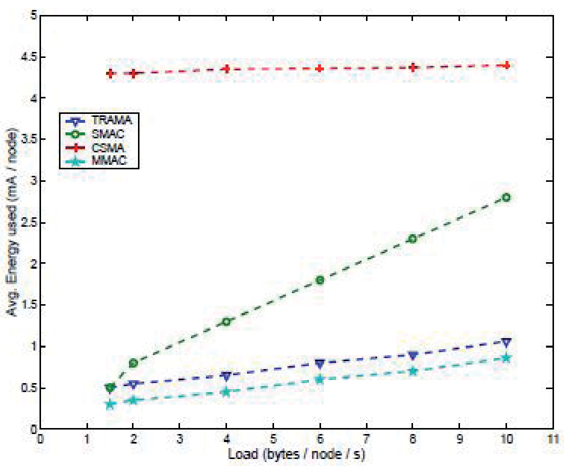


Figure 14. Average energy consumed per node (variable traffic) [26]

4. **FlexiMAC** [27]: copes with some network dynamics and node mobility. The first phase of the protocol aims to build a data gathering tree and nodes'schedules which is maintained throughout their lifetime in the network. CSMA/CA is used during this phase for nodes transmission. After this phase of the protocol, nodes perform regular data gathering tasks using their TDMA schedules. They also can modify their schedules when the network topology changes. FlexiMAC ensures the following functionalities:
   - Data Gathering Tree Construction: for routing data from a source node to the base station through a tree-path and to collect node's first-level neighbor IDs. To minimize energy consumption, nodes select the closest node within a default transmission radius of their parents and using the minimum transmission power needed to reach their parents. The data gathering tree helps the sensor nodes to know the local network topology, thus each node knows its parent, children, descendants, and first-level neighbors.
   - Time slot assignment: the slot distribution follows a depth first search (DFS) which allows data sent by a source node to be forwarded by routers to the base station in an

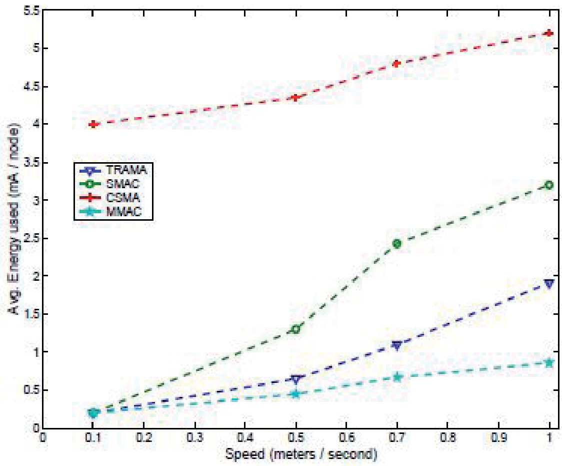


Figure 15. Average energy consumed per node (increasing mobility) [26]

interleaving manner. The slot numbering starts with number 2 because slot number 1 is reserved for management of network dynamics; hence the name (FTS: Fault Tolerant Slot). During this short FTS period, all nodes in the network are in the listen mode. The FTS is a large slot in which access is based on a CSMA mechanism. The main objective of the FTS period is to handle the mobility of the nodes and topology changes (orphan nodes or new nodes use this slot to ask for communication slots). Each node in the network has three modes; Transmission mode, reception mode and the sleep mode. Nodes switch to the transmission mode for scheduled Transmit Slot List (TSL), and receive mode for their scheduled Receive Slot List (RSL). Apart of these three specific slots, each node uses two other slots. The first one is used for local time synchronization and local repair and is called Multi-Function Slot (MFS). The second slot is called the Conflict Slot List (CSL) and is used to record slots that are used by a node′s first-level and second-level neighbors. A node selects a slot only if it does not belongs to its RSL, TSL, and CSL. Once the slot is selected, the node propagates it to its first-level neighbors. Each of these nodes propagates the selected slot to its direct neighbors (second-level neighbors). This approach avoids the multiple slot selection by different nodes, thus FlexiMAC guarantees a collision-free traffic.

The main issue of FlexiMAC is time synchronization. To solve this problem the protocol performs time synchronization locally, thus each node broadcasts its clock and the current global highest slot number known to that node to its children. Furthermore, robustness and optimality of the tree structure represent the main drawback of this protocol. In fact, when a link fails, a tree reconstruction even localized is necessary. Finally, the first phase where data gathering tree and nodes’schedules are built, consumes a large amount of energy.

6. MAC protocols in real-time wireless sensor networks

In real-time systems, correctness of the computations depends on their logical correctness and on the time at which the result is produced. Real-time applications become a very

Protocol Name	Scheme Used	Advantages	Disadvantages
MS-MAC [25]	it varies the sleeping time dynamically according to the velocity of the mobile node	performs better than S-MAC [3], has adaptive sleep period and overcomes the problem of mobility of nodes from one cluster to another	cannot ensure reliable communication between stationary nodes and mobile nodes
MMAC [26]	dynamically adapts to changes in mobility patterns by introducing a mobility-adaptive frame time	handles weak and strong mobility	requires the knowledge of node position
Flexi MAC [27]	defines a contention period in which nodes exchange packets to build a data-gathering tree rooted at the sink	able to cope with some network dynamics and node mobility	robustness and optimality of the tree structure represent the main drawback

Table 4. Summary of MAC protocols in mobile WSNs

important field of research as it covers an important domain of application including ABS, aircraft control, ticket reservation system at airport, over-temperature monitor in nuclear power station, mobile phone, etc. For many WSN applications like medical care and fire monitoring, real-time constraints must be considered as an important factor.

Real-time in WSNs can be divided into soft real-time (SRT) where just a portion of messages can arrive late, and hard real-time (HRT) where every message must arrive before its deadline and any deadline miss is considered as a failure of the system. As discussed in the previous sections there are three main categories of MAC protocols in WSNs, but till now we didn't discuss how much they can be adapted to the real-time application requirements in WSNs.

The schedule based protocols are not well adapted to real-time applications, because they are not good for event driven reporting, even if it is easier to define delay deadline at MAC layer with deterministic scheduling protocols. In general, compared to the contention based protocols, the delay is higher in TDMA based approaches. However, most of the discussed contention based protocols in the previous sections are not suitable for real-time applications. S-MAC [3] is a reference of contention based protocols, however it cannot be used for real-time application because a given node in the network has the possibility to win consecutively the contention many times while other nodes in the network are awaiting for the contention, consequently they risk a message delay and probably they miss their deadlines. Similarly to S-MAC, T-MAC proposed [5] is not suitable for real-time applications as well. because the first packet sent may reach the sink node with a very high delay.

During the last years, some researchers focused on developing protocols for real-time WSNs based on S-MAC. In the following, we will discuss some of these protocols and we summarize them in Table 5 at the end of this section.

1. **Virtual TDMA for Sensors (VTS) [28]**: this protocol is designed for soft real-time WSN applications and it is based on S-MAC [3]. However, unlike S-MAC, only one node can transmit in every listen/sleep cycle as shown in Figure 16, thus at each TDMA slot (a

frame in S-MAC), a given packet travels strictly one hop. The number of slots is equal to the number of nodes in a cell (cluster) and the nodes in a cluster will transmit in different time slots. VTS uses a special packet called control (CTL) packet for synchronization,

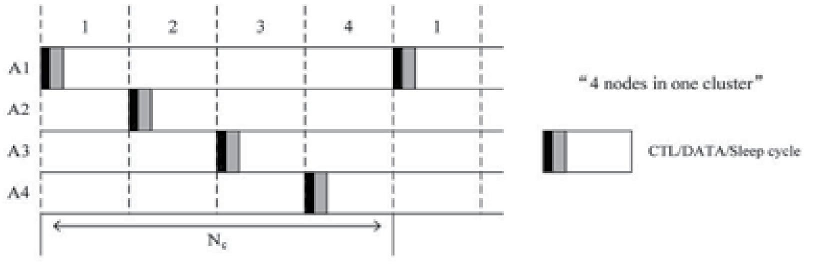


Figure 16. Structure of VTS TDMA frame [28]

schedule discovery, keep-alive beacon, new node discovery and channel reservation. The first phase of the protocol aims to setup the network exactly as the S-MAC synchronization mechanism. In the end of this phase, a virtual superframe of N_c timeslots is formed when all the nodes have sent their first CTL packet. To overcome the problem of topology change when some nodes join or leave the network, the nodes adjust the length of the superframe.

At the beginning of each time slot, all the nodes wake up and listen. The owner of the timeslot performs a carrier sense and broadcasts the CTL. To transmit data, VTS uses CSMA/CA mechanism, and there are three kind of transmission: unicast packet transmission, broadcast packet transmission and no data transmission.

In unicast packet transmission, the owner of the time slot sends a CTL$_{RTS}$ and the other nodes go to the sleep state. When the sequence of DATA/ACK are received, the transmission is finished and both nodes go to sleep mode.

In broadcast packet transmission, the owner of the time slot sends a CTL$_{BCAST}$. Unlike unicast packet transmission, the node keeps listening because the destination is a broadcast address. Without waiting for any CTS reply, a sender can send the broadcast packet. After receiving the packet, nodes go to sleep without sending an ACK.

In no data transmission, a CTL$_{BCAST}$ is sent. Nodes adjust the clock reference, clear sender inactivity counter and go to sleep.

With respect to S-MAC [3], VTS decreases energy consumption and the latency of packet transmission only when there are a few nodes, but the energy consumption increases when the number of nodes becomes important.

2. **Novel real-time MAC layer protocol [29]**: this protocol is also based on S-MAC [3] and it is designed for soft real-time applications in WSNs. The novel real-time MAC protocol is intended for single stream communication. It means that there is only one source and one sink during the lifetime of a communication stream in a randomly deployed WSN. Similarly to VTS [28], the novel real-time MAC layer protocol uses a control packet called Clear Channel (CC) to assign an appropriate value to the Clear Channel Flag (CCF) of every sensor node.

   The node that has a CCF value equal to '1' (the initialized value of all nodes) can transmit as well as receive data packets, while it can only receive if its CCF value is '0'. CC control packet has a Clear Channel Counter (CCC) from '0' to '3'. This value decreases by one with one hop transmission of CC, thus it is equal to '3' at the originating node. Furthermore, when the control packet CC has a value equal to '2' or '3' in a node, then CCF of that node will remain '0', otherwise CCF of that node will become '1', which means that it can now initiate a data packet transmission. The novel real-time MAC layer protocol reduces the latency with respect to S-MAC. However, the overhead is higher due to CC control packet. In addition, the novel protocol is not suitable to multi-streams communication.

3. **Low-power real-time protocol (LPRT) [30]**: is a hybrid solution designed for soft real-time applications in WSNs. The network is considered as a star topology where all the sensors communicate with the base station (the coordinator). In this case, the only way to extend the network range is to add other base stations.

   Figure 17 shows the superframe of LPRT where the communication starts from the base station and the frame is divided into several mini-slots (the number of mini-slots is known in advance). The beacon frame noted B in Figure 17 is followed by a contention

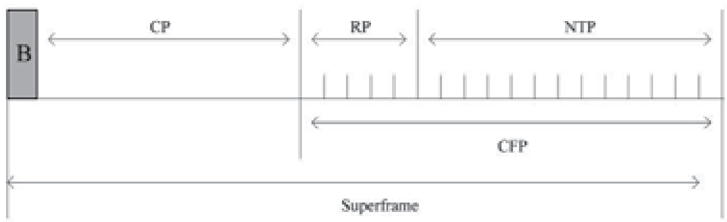


Figure 17. Superframe structure for the LPRT protocol [30]

   period (CP) where any station can transmit its packets using CSMA/CA. The contention free period (CFP) is placed after the CP period and it is divided into two parts: an optional retransmission period (RP) and a normal transmission period (NTP). During the CFP period, it is possible to transmit non-real-time asynchronous traffic if it cannot be completed before the beginning of the CFP. The transmission during CFP is scheduled by the base station based on a resource grant (RG) field of the beacon frame . The structure of the beacon frame is shown in Figure 18. The superframe duration field gives the duration of the current superframe in multiples of a minimum superframe duration time. The RG list field starts with an RG list length RGLL which specifies the quantity of allowed resource grant (RG). Each resource grant in the RG list field is expressed by a transmission direction (TD) bit to identify either the transmission is an uplink or a downlink to/from a device identified by the association ID (AID) field and the initial transmission slot (ITS) field used to indicate the end of the RG and the begin of the next RG in the RG list.

   The ACK list is composed by an ACK length (AL) field and an ACK bitmap field containing one bit for each uplink RG of the previous superframe. A successful transmission is indicated by a 1 in the respective bitmap position, while a lost or corrupted transmission is indicated by a 0.

   The main drawback of this protocol is the difficulty to handle a large multi-hop wireless sensor network and other network topologies.

4. **SPEED-MAC [31]**: the Speedy and Energy Efficient Data Delivery MAC Protocol aims to minimize the message delivery latency and the energy consumption during the

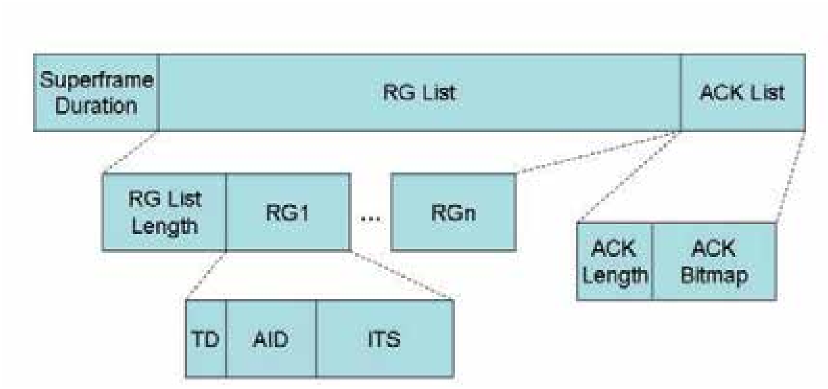


Figure 18. Structure of the beacon frame payload [30]

idle period. To meet these goals, SPEED-MAC divides the cycle time into an event announcement period (called also the signalling wakeup period), and a data transmission period as shown in Figure 19. During the first period, SPEED-MAC sends a control signal called a SIGNAL packet before sending any data packet to notify the occurrence of an event. This technique is used to detect a collision for multi-source events and also to schedule the wakeup during the data transmission period. In addition, SPEED-MAC uses different wakeup strategies for single-source events and multi-source events. One of these strategies is called *Adaptative Wakeup* for demand-driven data transmission, thus if there is no traffic, nodes wake up at only event announcement periods, otherwise they additionally wake up for data transmission periods. The SIGNAL packet contains the sender address

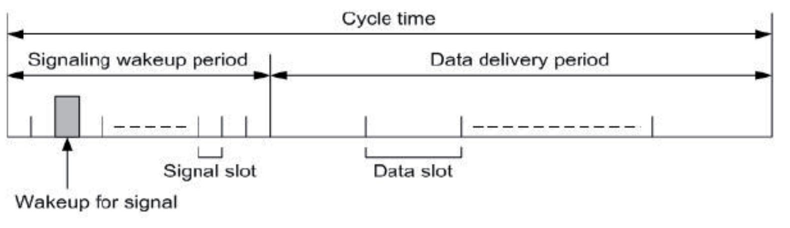


Figure 19. Cycle time breakdown in SPEED-MAC [31]

and a single bit for collision detection. On wakeup, a node receives the SIGNAL packet only if the sender is one of its children. For this purpose, each node needs to maintain its children list during the network initialization phase. The collision bit is used not only for collision detection but to determine the nature of the event as well (single-source event or multi-source event). All the senders overhear the signal transmission from their parents to check whether the data transmission period is for a single-source event or for a multi-source event.

During the Data transmission period, the packet consists of DATA and several flags. For a single-source event, the period consists of DATA and ACK, otherwise, if the event is a multi-source one, then the period consists of RTS/CTS/DATA/ACK with CSMA/CA.

Therefore, SPEED-MAC uses contention-based data transmission for multi-source events and fixed scheduled transmission for single-source events.

The total number of signal slots (data slots) in the event announcement period is determined by the depth of the routing tree. Note that the number of data slots is the same as the total number of signal slots (see Figure 19).

Figure 20 (a) shows that SPEED-MAC has much lower energy consumption than SMAC and DMAC. We can also see in Figure 20 (b) that SPEED-MAC outperforms the other protocols even when the total number of source nodes increase because the protocol reduces the idle listening overhead and removes unnecessary wakeups.

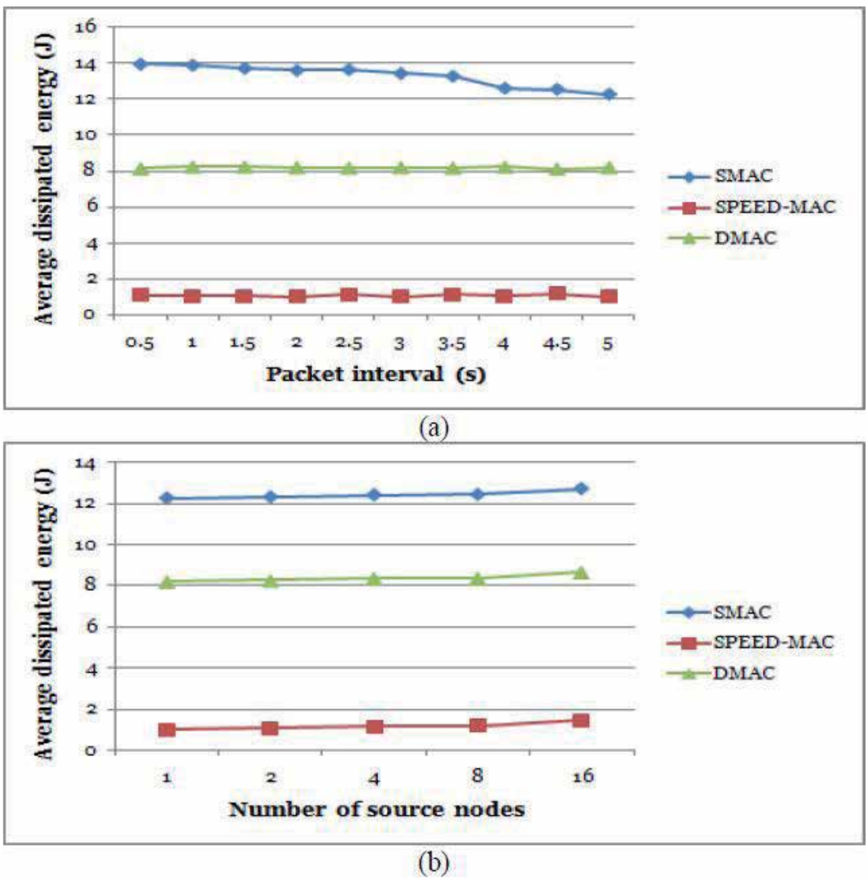


Figure 20. Average dissipated energy by varying (a) the packet generation interval and (b) the number of source nodes [31]

5. **TOMAC [32]**: the real-Time message Ordering in wireless sensor networks using the MAC layer (TOMAC) of [32] is a hard real-time layer protocol for WSNs. To access the channel, the protocol uses the measure of elapsed waiting time of a message as an input parameter. Therefore, the node having older information to send get the highest priority, thus, it access the channel first. This mechanism guarantees that events are transmitted in the order of

their occurrence in the real world. This message ordering is designed for one-hop distance mesh topology.

TOMAC assigns a priority to each produced message. On the MAC layer, the generated packet gets the minimum priority value that keeps increasing with a given gradient e.g $\frac{d_{ri}}{dt} = \frac{1}{\mu_s}$ while the packet is not transmitted (due to a sleep time in a low duty cycle protocol or due to a busy channel) or the maximum priority value ri_{max} is reached (ri is the priority of message i). Figure 21 shows how the message priority increases for the waiting packet of nodes A,B, and C. It is possible to assign a priority value ri before proceeding the

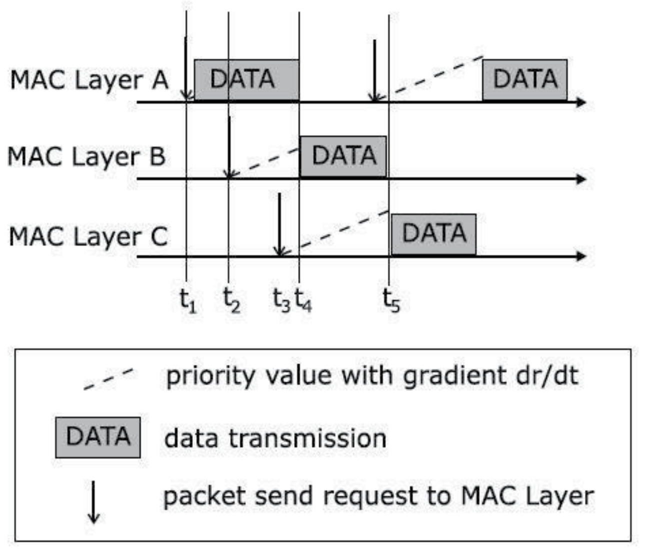


Figure 21. Increasing priorities for waiting packets [32]

message to the MAC layer. This exception may happen when the notification of sensor events should be transported to a destination in a time-ordered manner.

For WSNs, the time resolution is an important aspect, thus the gradient of priority must be relatively high. Therefore, if the maximum possible value of a priority r_{max} is very high, a very high number of priority steps have to be distinguished. The following function is used to determine r_{max}:

$$r_{max} = \Delta_t \; . \; \frac{d_r}{dt} \quad [steps]$$

Δ_t describes the maximum total possible delay until the transmission of a waiting packet.

To handle a high number of priority steps ($> 10^6$) like in low duty cycle TDMA protocols, TOMAC uses a non-destructive bit wise arbitration (NDBA). NDBA is implemented in the MAC layer of the Controller Area Network (CAN-bus) [33]. Each message is assigned an ID number, thus the message having the lowest ID accesses the channel. The bits of the message-ID are interpreted as dominant (0) and recessive (1) and are sent into the medium by the attached nodes. If a recessive bit is sent by a sensor node, this one loses the channel access. In a WSN context, the ID is replaced by the binary complement of the priority, then NDBA is used to solve concurrent access for the node with highest priority.

Protocol Name	Scheme Used	Advantages	Disadvantages
Virtual TDMA for sensors [28]	a node can transmit in every listen/sleep cycle	for WSNs with few nodes, this protocol decreases energy consumption and the latency of packet transmission	the energy consumption increases when the number of nodes becomes important.
Novel real-time MAC layer protocol [29]	it uses a control packet called Clear Channel (CC) to assign an appropriate value to Clear Channel Flag (CCF) of every sensor node	reduces the latency with respect to S-MAC [3]	the overhead is higher due to CC control packet
Low-power real-time protocol [30]	all the sensors communicate with the base station and a communication starts from the base station. The frame is divided into several mini-slots.	it provides low latency and it is flexible	it can not handle a large multi-hop wireless sensor network and other network topologies
Speed MAC [31]	divides the cycle time into an event announcement period, and a data transmission period	minimizes the message delivery latency and the energy consumption and also distinguishes single-source events from multi-source events	no guarantee is given as the congestion can not be predicted. Packet loss. Overflow deadline is always possible.
TOMAC [32]	to access the channel the protocol uses the measure of elapsed waiting time of a message as an input parameter	is a hard real time protocol	the protocol is designed only for one-hop distance mesh topology.

Table 5. Summary of MAC protocols in real-time WSNs

7. Conclusions

Throughout this document, we analyze the main MAC protocols designed for wireless sensor networks. We considered the technique being used and the problem that the protocols try to solve. Up to now, there is no MAC protocol considered as a standard, because MAC protocols are generally application specific.

The first category of MAC protocols treated in this survey is the contention based protocols, which suffer from packet collision. However they guarantee low latency and high throughput.

Unlike contention based protocols, the scheduled protocols have no problems with collision, but the synchronization is critical and these kind of protocols are not adapted to topology change. Some hybrid solutions are presented in this survey, above solutions are combined to achieve high performance under variable traffic patterns.

The second part of this document presents some MAC solutions where mobile nodes are implemented in the network. Finally many solutions are designed to deal with real time applications in WSNs.

In summary WSN MAC protocols are application dependent and focus on energy consumption and transmission latency, thus there is no protocols that consider all of the requirements of sensor networks.

Author details

Besem ABID, Hamida SEBA
Laboratoire GAMA, Université de Lyon, Université Claude Bernard Lyon1, Département Informatique, 01000 Bourg en Bresse, France

Souleymane M'BENGUE
R&D Water/Heat Software Department, ITRON International, Macon ,France

8. References

[1] IEEE. *IEEE Standard 802.11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1999.
[2] N. Abramson. The aloha system - another alternative for computer communication. In *Fall Joint Computer Conference*, 1970.
[3] Wei Ye, John S. Heidemann, and Deborah Estrin. An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In *IEEE INFOCOM*, volume 3, 2002.
[4] Suresh Singh and C. S. Raghavendra. PAMASŮpower aware multi-access protocol with signalling for ad hoc networks. *Computer Communication Review*, 28:5–26, 1998.
[5] T. Van Dam and K. Langendoan. Energy-efficient MAC: An adaptive energy-efficient MAC protocol for wireless sensor networks. In *Conference On Embedded Networked Sensor Systems*, 2003.
[6] Paul J. M. Havinga and Gerard J. M. Smit. E2MaC: an Energy Efficient MAC protocol for multimedia traffic. 1998.
[7] Jinsuk Pak, Jeongho Son, and Kijun Han. *A MAC Protocol Using Separate Wakeup Slots for Sensor Network*. 2006.
[8] Wei Ye, John S. Heidemann, and Deborah Estrin. Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE/ACM Transactions on Networking*, 12:493–506, 2004.
[9] J. Hill J. Polastre and D. Culler. Versatile low power media access for wireless sensor networks. In *Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Baltimore, 2004.
[10] Peng Lin, Chunming Qiao, and Xin Wang. Medium access control with a dynamic duty cycle for sensor networks. *Revue Francophone Des Laboratoires*, 3, 2004.
[11] Shih-Hsien Yang, Hung-Wei Tseng, Eric Hsiao kuang Wu, and Gen-Huey Chen. Utilization based duty cycle tuning MAC protocol for wireless sensor networks. In *Global Telecommunications Conference, . GLOBECOM . IEEE*, 2005.
[12] Y. E. Sagduyu and A. Ephremides. The problem of medium access control in wireless sensor networks. *IEEE Wireless Communications*, 11:44–53, 2004.

[13] Avinash Sridharan and Bhaskar Krishnamachari. Max-min fair collision-free scheduling for wireless sensor networks. In *Power Conversion Conference*, 2004.

[14] Gang Lu, Bhaskar Krishnamachari, and Cauligi S. Raghavendra. An Adaptive Energy-Efficient and Low-Latency MAC for Data Gathering in Wireless Sensor Networks. In *International Parallel and Distributed Processing Symposium/International Parallel Processing Symposium*, 2004.

[15] Venkatesh Rajendran, Katia Obraczka, and J. J. Garcia-Luna-Aceves. Energy-efficient collision-free medium access control for wireless sensor networks. In *Conference On Embedded Networked Sensor Systems*, pages 181–192, 2003.

[16] Khaled A. Arisha, Moustafa A. Youssef, and Mohamed F. Younis. Energy-aware tdma-based mac for sensor networks. In *IEEE Workshop on Integrated Management of Power Aware Communications, Computing and Networking (IMPACCT 2002)*, 2002.

[17] Sinem Coleri Ergen and Pravin Varaiya. PEDAMACS: Power Efficient and Delay Aware Medium Access Protocol for Sensor Networks. *IEEE Transactions on Mobile Computing*, 5:920–930, 2006.

[18] Michael I. Brownfield, Kaveh Mehrjoo, Almohonad S. Fayez, and Nathaniel J. Davis. Wireless sensor network energy-adaptive mac protocol. In *CCNC IEEE Consumer Communications and Networking Conference*, 2006.

[19] André M. Barroso, Utz Roedig, and Cormac J. Sreenan. ţ-MAC: an energy-efficient medium access control for wireless sensor networks. In *European Workshop on Wireless Sensor Networks*, pages 70–80, 2005.

[20] IEEE. *IEEE 802.15.4. Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, 2003.

[21] Injong Rhee, Ajit Warrier, Mahesh Aia, and Jeongki Min. Z-MAC: a hybrid MAC for wireless sensor networks. In *Conference On Embedded Networked Sensor Systems*, pages 90–101, 2005.

[22] Injong Rhee, Ajit Warrier, Jeongki Min, and Lisong Xu. DRAND: : distributed randomized TDMA scheduling for wireless ad-hoc networks. In *Mobile Ad Hoc Networking and Computing*, pages 190–201, 2006.

[23] Gahng-Seop Ahn, Se Gi Hong, Emiliano Miluzzo, Andrew T. Campbell, and Francesca Cuomo. Funneling-MAC: a localized, sink-oriented MAC for boosting fidelity in sensor networks. In *Conference On Embedded Networked Sensor Systems*, pages 293–306, 2006.

[24] Katayoun Sohrabi, Jay Gao, Vishal Ailawadhi, and Gregory J. Pottie. Protocols for self-organization of a wireless sensor network. *IEEE Wireless Communications*, 7:16–27, 2000.

[25] Huan Pham and Sanjay Jha. An adaptive mobility-aware MAC protocol for sensor networks (MS-MAC). In *IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems*, 2004.

[26] Muneeb Ali, T. Suleman, and Zartash Afzal Uzmi. MMAC: a mobility-adaptive, collision-free MAC protocol for wireless sensor networks. In *Power Conversion Conference*, pages 401–407, 2005.

[27] Winnie Louis Lee, Amitava Datta, and Rachel Cardell-Oliver. FlexiMAC: A flexible TDMA-based MAC protocol for fault-tolerant and energy-efficient wireless sensor networks. In *IEEE International Conference on Networks*, volume 2, pages 1–6, 2006.

[28] Esteban Egea-lópez, Javier Vales-alonso, Alejandro S. Martínez-sala, Joan García-haro, Pablo Pavón-mariño, and M. Victoria Bueno-delgado. *A Real-Time MAC Protocol for Wireless Sensor Networks: Virtual TDMA for Sensors (VTS)*. 2006.

[29] Brajendra Kumar Singh and Kemal Ertugrul Tepe. A novel real-time MAC layer protocol for wireless sensor network applications. In *IEEE International Conference on Electro/Information Technology*, pages 338–343, 2009.

[30] J. A. Afonso, L. A. Rocha, H. R. Silva, and J. H. Correia. MAC Protocol for Low-Power Real-Time Wireless Sensing and Actuation. 10(13):1248–1251, December 2006.

[31] Sang Hoon Lee Lynn Choi. Speed-mac: Speedy and energy efficient data delivery mac protocol for real-time sensor network applications. In *Proceedings of the International Conference on Communications (ICC 2010)*, South Africa, May 2010.

[32] Albert Krohn, Michael Beigl, Christian Decker, and Tobias Zimmer. Tomac- real-time message ordering in wireless sensor net-works using the mac layer. In *Proceedings of 2nd In-ternational Workshop on Networked Sensing Systems*, pages 27–28, San Diego, June 2005.

[33] ISO. *ISO 11898-1. Road vehicles controller area network (can) part 1: Data link layer and physical signalling*, 2003.

Chapter 5

Bandwidth and Energy Consumption Tradeoff for IEEE 802.15.4 in Multihop Topologies

Abdeddaim Nazim, Darties Benoît and Theoleyre Fabrice

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48295

1. Introduction

With the development of electronic technology and embedded systems, wireless sensor networks (WSN) are gaining growing interest over the past few years. A WSN is a collection of autonomous devices organized into a cooperative radio network . Each node is connected to one or several sensors, and may monitor various physical or environmental conditions, such as temperature, vibration, motion, radiation activity. WSN can vary from a few to thousands nodes equipped with transceivers for low range wireless communication.

WSN benefit from low-cost technology, flexible deployment and self-management features, making them a serious alternative to traditional networks in many domains. WSN application fields are numerous and highly anticipated in the near future: e.g. structure health monitoring (SHM), habitat monitoring, localization and tracking, environmental sensing, industry, healthcare monitoring. Over the past 10 years, numerous projects have lead to real-life deployments : one emblematic example for SHM is operated on the Golden Gate bridge [1], which is monitored thanks to a 46 hops-network of 64 nodes equipped with accelerometers and temperature sensors. Collected data help, among other things, to determine the response of the structure to ambient conditions, to measure the structural accelerations from wind load or strong shaking from a potential earthquake. WSN may also operate in various productivity domains including agriculture. Camalie Vineyards deployed a smart infrastructure [2], leading to significant productivity gain. WSN-based approaches are also particularly relevant in extreme environmental conditions, i.e. for monitoring volcanic activity. Others real-life experimentations may include habitat monitoring (e.g. Great Duck Island project), wildlife tracking (e.g. Zebranet Project), healthcare monitoring (e.g. Mobihealth), etc.

When wireless sensor networks are designed on embedded systems for long term operations, energy conservation is an essential requirement. To achieve low-cost, low-speed ubiquitous communication between devices, IEEE 802.15 working group has proposed the standard IEEE 802.15.4: this standard specifies the physical layer low-rate energy-efficient transmissions, operating energy conservation though the use of sleep/wakeup scheduling protocols, thus minimizing the activity on sensor nodes. These features make IEEE 802.15.4 one of the main standards for low data rate wireless sensor networks. However, energy limitations have a negative impact on network capacities since transmissions are unavailable when nodes are in power-saving mode. In this chapter, we investigate the problem of increasing the capacity of IEEE 802.15.4 single-hop and multihop networks subjects to energy limitations. In the presented analysis, we assume classical scenarios where collected data must be routed to a unique collect point, i.e. a *convergecast* traffic pattern. The wide majority of IEEE 802.15.4 deployments also use a sink to collect data.

The chapter will be organized as follows. In Section 2, we analyze IEEE 802.15.4 mechanisms including node organization, MAC mechanisms, energy conservation, topology construction and node association. In Section 3, we detail how we should modify IEEE 802.15.4 to cope efficiently with multihop topologies, scheduling the transmissions. In Section 4, we quantify the impact of the cluster-tree algorithm on the network performances. We expose how the overall throughput can be improved with a novel cluster-tree construction algorithm defined formally as a Mixed Integer Linear Programming formulation. In Section 5, we quantify the impact of each parameter on the performances of IEEE 802.15.4. In particular, we present a self-configuration algorithm to dynamically adjust the Backoff Exponent so that the protocol always operates in optimal conditions. We finally conclude this chapter by discussing open challenges in this research area.

2. Background on IEEE 802.15.4

An IEEE 802.15.4 network comprises one PAN coordinator and a set of devices. Theses low-power devices are characterized by a limited transmission range and a limited quantity of energy. The IEEE has proposed the IEEE 802.15.4 to govern the medium access in this kind of networks, presented in [3]. The protocol uses a PAN coordinator, inter-connecting the WSN to e.g. the Internet. Although IEEE 802.15.4 was originally designed for single hop networks, the working group has proposed topologies to cope with multihop applications (Figure 1):

star: the PAN coordinator is in the radio range of all other nodes (i.e. each node forms a *branch of the star*). Single hop transmissions are in this case sufficient;

peer-to-peer: a node may communicate with any neighbor, the structure being decentralized. A routing protocol may enable multihop communications, using P2P transmissions at the MAC layer;

cluster-tree: a tree is constructed, rooted at the PAN coordinator. All the non leaf-nodes are designated as *coordinators* since they may forward the traffic to or from the root.

2.1. MAC mechanisms

In these topologies, IEEE 802.15.4 may work either in *non-beacon* or in *beacon-enabled* mode. The MAC strategy impacts the duty-cycle and thus the capacity and energy consumption. In IEEE 802.15.4, one set of nodes (the *coordinators*) regulates the transmissions. Any exchange is initiated from the non-coordinator to the coordinator. In particular, the coordinator buffers all the packets destined to others. Then, each non-coordinator MUST periodically ask for the buffered packets to the coordinator by sending a `data-request` — at most every `macTransactionPersistenceTime` since packets are removed from the buffer of

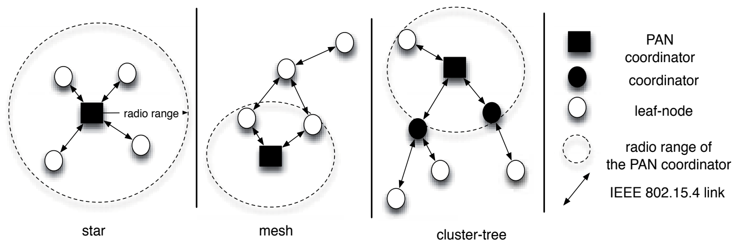


Figure 1. The different topologies proposed in IEEE 802.15.4

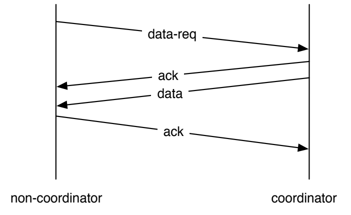


Figure 2. Indirect transmission (non-coordinator → coordinator)

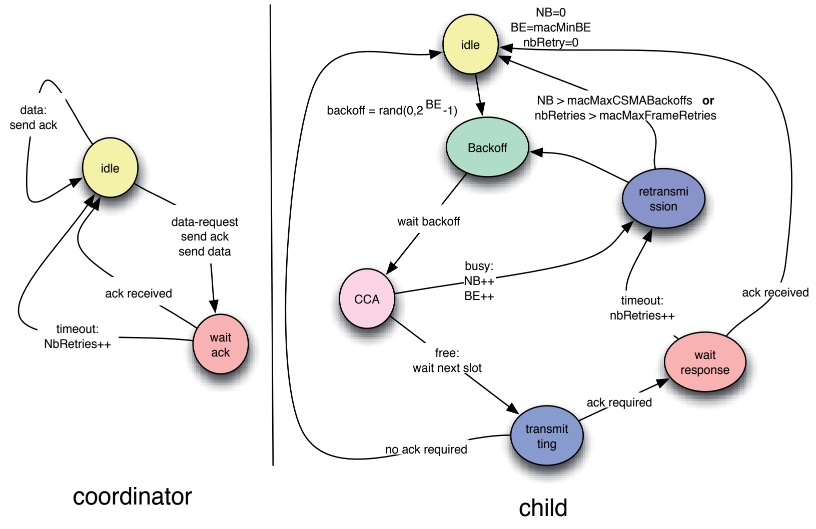


Figure 3. MAC procedure in the non-beacon mode

the coordinator after this timeout. IEEE 802.15.4 implements two *modes* for the medium access. In non-beacon mode, transmissions use a classical CSMA-CA scheme while the beacon-enabled mode introduces the concept of superframes to schedule the transmissions more appropriately and to assign dedicated timeslots.

2.1.1. Non-beacon mode

In non-beacon mode, a node just uses a classical CSMA-CA procedure to transmit its packets (Figure 3)

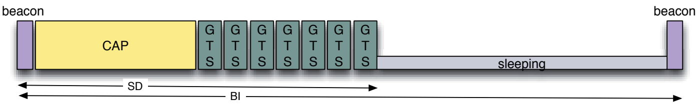


Figure 4. The superframe structure in IEEE 802.15.4

1. The node is in the idle state, and initializes the value of the backoff exponent (BE).
2. It chooses a random backoff comprised between 0 and $2^{BE} - 1$;
3. After this backoff, the node must trigger a Clear Channel Assessment (CCA): if the channel is free, the transmission begins. Else, the node triggers a new backoff after having incremented the Backoff Exponent (BE) and the number of backoff retries (NB). If NB exceeds the threshold `macMax CSMABackoffs`, the node comes back to the idle state;
4. After having transmitted the packet, the node comes back to the idle state if it does not need acknowledgement. Else, it waits for the ack;
5. If the ack is correctly received, the node considers the transmission successful and proceeds to the next packet. Else, it increases the number of retransmissions and tries a new retransmission if is does not exceed the threshold value `macMaxFrameRetries`.

Figure 3 presents a simplified state diagram for the unslotted mode, separated for coordinators and children. As coordinator, the node has just to wait solicitation: the medium is pre-reserved by the other side. As child, it must choose a random backoff, trigger a CCA after the backoff is finished, and then transmit the frame if the medium is idle. The number of backoff retries and retransmissions are upper-bounded respectively by `macMaxCSMABackoffs` and `macMaxFrameRetries`. Besides, a node must ensure the `BE` values does not exceed `macMaxBE`. A child may wait for a response from the coordinator: either a simple `ack` if it transmitted a `data` packet or an `ack` followed by a `data` packet if it transmitted a `data-request`. As coordinator, the only response to wait can be an `ack`. IEEE 802.15.4 also proposes a Battery Life Extension (BLE) option for very energy-constrained devices. With this option enabled, a node will choose a minimal Backoff Exponent value after a successful transmission. Thus, it will gain medium access quicker than other nodes without BLE.

2.1.2. Beacon-enabled mode

In *beacon-enabled* mode, IEEE 802.15.4 introduces the concept of superframes (Figure 4). Each coordinators sends periodically – every `Beacon Interval` (BI) – a `beacon`, piggybacking control information. Just after having received the `beacon`, children may engage a transmission during the Contention Access Period (CAP) using slotted CSMA-CA. Children may in particular map a reservation during the CAP to obtain a dedicated Guaranteed Timeslot (GTS) placed just after the CAP. Finally, the children and the coordinator may sleep until the next beacon. The whole active part of the superframe lasts for a `Superframe Duration` (SD). When a node has finished to participate to the superframe, it may sleep until the next `beacon` reception/transmission. The Superframe Duration (resp. Beacon Interval) are defined through the Beacon Order (resp. Superframe Order) values, according to the following relation:

$$SD = aBaseSuperFrameDuration * 2^{SO} \quad (1)$$

$$BI = aBaseSuperFrameDuration * 2^{BO} \quad (2)$$

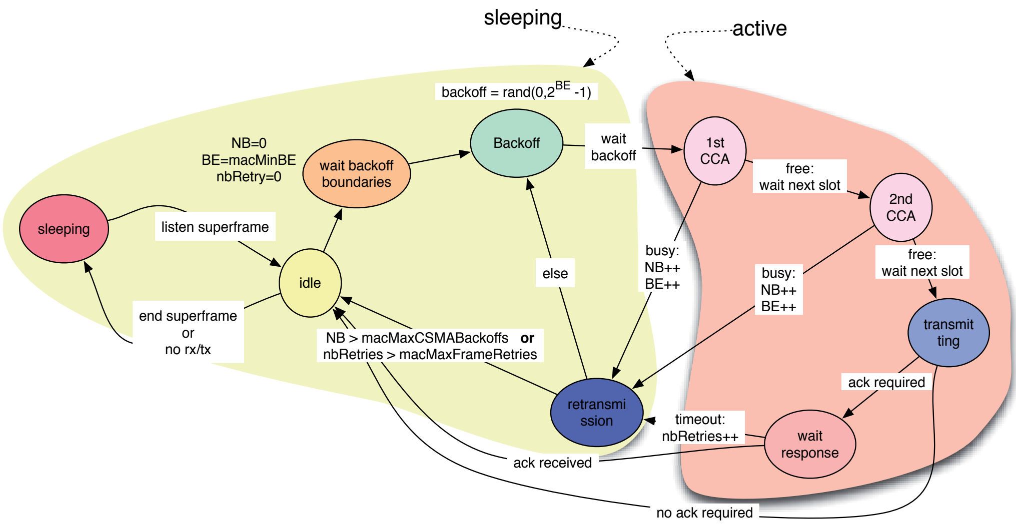


Figure 5. MAC procedure in the beacon-enabled mode for the child

By adjusting the BO and SO values, we can obtain a tradeoff between network capacity and energy savings. For instance, a duty cycle of 1% can be obtained if $SO = BO - 7$ ($2^{-7} < 1\%$). The beacon-enabled mode is only available in the star and cluster-tree topologies since we must maintain a hierarchy of coordinators to schedule appropriately the `beacons`. Figure 5 illustrates a simplified state diagram of the child mode in slotted mode. The coordinators implement the same behavior as depicted in Figure 3. In particular, the node triggers two CCA before transmitting a packet (the `CW` parameter in the IEEE 802.15.4 standard). Besides, a child may sleep all other the time except during CCA, the transmission and while waiting for a response from the coordinator (a unique `ack` or a pair of `ack` and `data` packets).

2.2. Energy savings

2.2.1. Peer-to-peer topology

If no energy optimization technique is implemented in a peer-to-peer topology, a Full-Function Device (FFD) may receive a packet at any time from a neighbor. To save energy, the MAC layer must implement a synchronization mechanism so that two neighbors can agree on a sleeping schedule. The pair of nodes must be periodical awake to exchange packets. A Low-Power-Listening protocol with preambles may be implemented to be sure the receiver is awake. This technique was proposed to reduce the energy consumption in very low traffic conditions [4]. When a node wakes-up and senses a signal, it must stay awake to receive further a `data` packet. Bachir *et al.* [5] proposed to piggyback information in the preamble so that a node knows it is the destination of the next packet when it receives the preamble, technique known as *preamble sampling*. The nodes may also agree on their wake-up schedule, as proposed in S-MAC [6]. Nodes broadcast periodically `beacons` to publish their schedule and synchronize with their neighbors. This mechanism may be coupled with the preamble to deal with clock drifts and reduce the synchronization requirements, as in [7]. However, all these propositions are outside the scope of the standard and to the best of our knowledge were never evaluated with IEEE 802.15.4.

2.2.2. Star topology

In single hop networks (i.e. star topology), all transmissions are initiated by the *followers* (i.e. nodes which follow the PAN coordinator). Thus, a node may sleep safely, asking periodically for pending packets in the PAN coordinator with a `data-request`. Nodes may reduce their energy consumption: only the PAN coordinator has to stay awake continuously.

2.2.3. Cluster-trees

non-beacon mode In a cluster-tree, we may authorize some nodes to be categorized as Reduced-Function Devices (RFD). These nodes cannot relay packets because of energy constraints. Since they would constitute leaf nodes in the cluster-tree, they may sleep safely. However, all other nodes are Full-Function Devices (FFD) and cannot sleep: energy savings are very limited.

beacon-enabled mode As highlighted previously, all the transmissions are initiated by the children, i.e. the coordinator cannot start a transmission before a solicitation of its child. This feature permits to implement efficient power-saving mechanisms. As *follower* (i.e. a node which participates to a superframe without coordinating it), a node must wake-up to receive at least one `beacon` every `macTransactionPersistenceTime`. If packets are pending, it must retrieve them immediately by transmitting a `data-request`. As soon as a follower has neither pending packet nor packet in its own buffer to transmit to its parent, it may sleep. As coordinator, a node must stay awake during the whole active part of its superframe. For all these reasons, the beacon-enabled mode should be privileged since this constitutes the only way to optimize the energy consumption in multihop topologies.

2.3. Topology construction

Except for the peer-to-peer topology where a node can communicate with any neighbor, the cluster-tree and star topologies require a node to be associated before transmitting packets. An unassociated node must discover a neighboring associated coordinator and send an `association-request`. Then, the node sends a `data-request` to retrieve an `association-reply`. If the newly associated node is a Full-Function-Device (FFD), it may accept now the association requests. However, the standard does not specify what coordinator a node should choose to associate with. The properties of a cluster-tree when a node associates to the first available parent is studied in [8]. The impact of mobility on the topology formation process was also investigated in [9]: it creates some convergence problems. Zigbee [10] proposes to couple the association procedure with an address assignment scheme: addresses being hierarchical, routing is simplified. However, the limited pool of addresses could create the orphan problem, as highlighted in [11]. Thus, the cluster-tree must be well balanced. Only few attention has been given on determining the characteristics of the cluster-tree should own and what algorithms could obtain them.

2.3.1. Active versus passive discovery

To discover an already associated coordinator, we may implement two strategies:

active discovery: a node enters in *active scan* and sends a `beacon-request` on each operational channel. An already associated coordinator MUST reply with a `beacon` if the PAN work in the *non-beacon* mode, otherwise the coordinator ingnore the

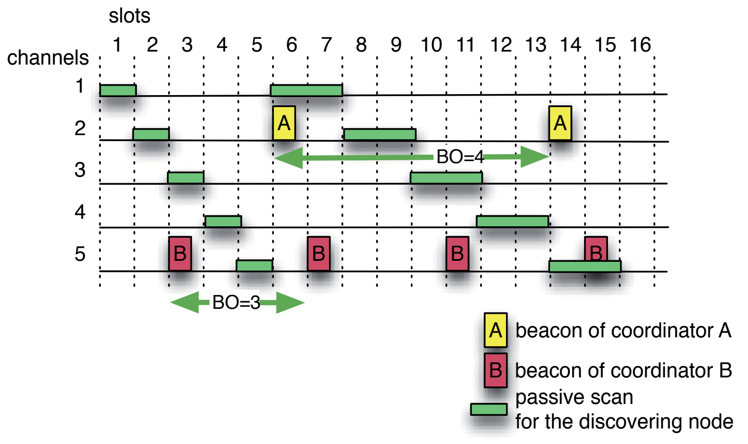


Figure 6. Passive Discovery in Multichannel IEEE 802.15.4

beacon-request and continue sending its periodic beacons. Upon reception of a beacon, the node can engage the association procedure .

If many nodes begin an active scan simultaneously, we may face to collisions. Besides, this method must be carefully implemented in beacon-enabled mode: a coordinator may sleep during the passive part of the superframe. Thus, one beacon-request has to be transmitted every Superframe Duration (SD), each channel being scanned during Beacon Interval (BI).

passive discovery: In beacon-enabled mode, a coordinator MUST transmit every Beacon Interval (BI) its beacon at the beginning of the active part of its superframe. Thus, a node may implement a *passive scan*: it has to stay at least BI on a channel to receive any beacon from already associated coordinators. If BI and channel are not known a priori, the node may assume the worst case, leading to very long discovery times. This strategy is impossible in non-beacon mode since an associated coordinator does not send any periodical packet.

The discovery process may be adapted for mobile nodes in beacon-enabled mode, as in [12] . They introduced a meeting channel: the manager in charge of mobile nodes send periodically their beacons on the meeting channel. Since the number of managers is limited and are geographically distributed, collisions may stay limited. Some focused on the discovery process for multichannel IEEE 802.15.4 (see [13]). The scan is optimized to discover first coordinators with a small Beacon-Interval (larger probability to be discovered first when a channel is scanned during a small duration). Figure 6 illustrates this behavior: the discovering node scans for small increasing durations each channel. Karowski *et al.* [14] extended this approach by reducing the redundancy in the discovery phase (a periodical slot has not to be scanned several times).

2.3.2. Required properties

Since Wireless Sensor Networks comprise a wide range of applications, we cannot focus on a particular topology: we may deploy a grid for surveillance applications, or pseudo-randomly

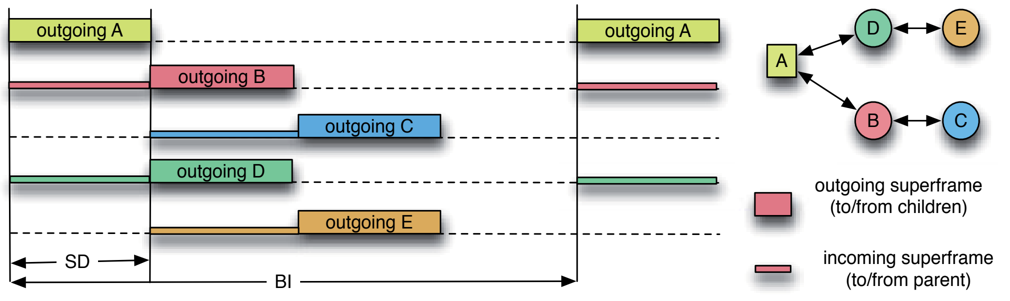


Figure 7. Scheduling of superframes in multihop topology.

for building automation. Moreover, the topology may be dynamic: we must be able to insert or remove nodes during the lifetime of the network. Thus, algorithms and protocols must distributed (decisions should use only local knowledge to decrease the overhead and improve the scalability) and self-configurable (the protocol must assign autonomously addresses and obtain its parameters values). In the same way, we must cope with different traffic patterns:

convergecast (Multipoint to Point) : all the packets are transmitted to one sink;

Point to Multipoint : information is flooded in the network;

Point to Point : direction communications may exist between e.g. a sensor and an actuator. For instance, Zigbee [10] proposes routing shortcut mechanisms to cope with this traffic pattern. On the contrary, RPL [15] uses only the DAG to route packets.

3. Cluster-tree scheduling in multihop topologies

Because we aim at minimizing the energy consumption in multihop topologies, we focus here on the cluster-tree topology in beacon-enabled mode. As highlighted in section 2.2, other modes do not permit to save energy easily.

3.1. Problem statement

Any device has to participate to the superframe of its parent, denoted *incoming superframe*. Additionally, a FFD has also to maintain its own superframe, called *outgoing superframe*. The standard mentions the outgoing and incoming superframes are interspaced by `StartTime`. Obviously, `StartTime` may be superior than Superframe Duration, else both superframes will overlap, creating collisions. Since cutting off its radio has an energy cost, we may choose `StartTime` $= SD$ so that the outgoing superframe follows directly the incoming superframe. Figure 7 illustrates a simple cluster-tree with this approach. Thus, all the coordinators with the same depth in the cluster-tree will start the active part of their superframes simultaneously. They will in particular send their `beacon` at the same time, creating a collision and making IEEE 802.15.4 inefficient. CSMA-CA performs very poorly in multihop topologies because of hidden nodes, unfairness and collisions, as demonstrated in [16]. Thus, we face to the same pathological cases in IEEE 802.15.4: collisions increase quickly with the number of nodes and the density, making the network unsuitable for large-scale applications. We clearly have to solve these problems to make IEEE 802.15.4 efficient in multihop topologies.

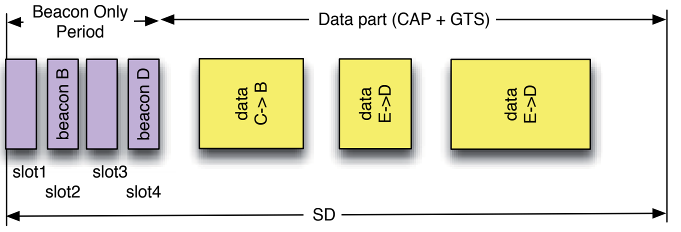


Figure 8. Beacon-Only Period to solve collisions among `beacons`

3.2. Beacon-only period

To reduce the number of collisions, we may reserve a *Beacon Only Period* (BOP) at the beginning of the superframe, as proposed by the IEEE 802.15.4b [17]. The BOP is divided into slots: each coordinator has to choose a BOP slot to transmit its `beacon`. Thus, several coordinators may share the same superframe without interference by choosing different BOP slots for their `beacons`. Let consider Figure 8 illustrating such approach (with the same topology as Figure 7). We only represented the superframe of the nodes B, C and D. We have in this case 4 dedicated timeslots used for the `beacons`: B chooses the second slot for its `beacons` while node D chooses the 4^{th} one. An algorithm must be implemented to detect collisions of `beacons` during a BOP slot. Indeed, a node N may detect a collision but the transmitters may not if e.g. N is placed in the middle of both coordinators. While this technique limits collisions between `beacons`, `data` may keep on colliding. Indeed, the different coordinators share the same Contention Access Period, creating hidden terminals. Unfortunately, this scenario occurs frequently, as demonstrated in [18].

3.3. Superframe scheduling

A second solutions consists in scheduling the superframe so that nodes which participate simultaneously to the superframe of two different coordinators do not interfere with each other, as advocated by IEEE 802.15.4b [17]. In this way, the remove entirely the collisions among nodes which participate to different superframes. The standard advocates all the nodes to use the same Beacon Order (BO) and Superframe Order (SO) values. In this case, we can construct a schedule of superframes using a Time Division Multiple Access (TDMA) approach. Each slot may contain exactly one superframe: a node has consequently to find a slot not used by an interfering coordinator to transmit its superframes. Using different SO and the same BO is equivalent to reserving several consecutive slots for one coordinator. However, this implicates the scheduling problem by inserting new constraints. In particular, the consecutive slots must start at a multiple of SD. This scheduling problem has received strong attention in the past. A clustering approach could simplify the scheduling as in [19], in which the PAN coordinators find an accurate scheduling while controlling the power of each superframe. However, this requires a centralized approach, based on information complicated to obtain (i.e. interferences are not trivial to estimate). This scheduling has been extended for the mesh topologies in [20]. Thus, such solution does not permit to save energy. A greedy localized solution was proposed in [21]: a node picks a free slot in its 2-neighborhood to schedule its superframe. However, it only takes into account interferences among coordinators, and not with children, which occur frequently in a cluster-tree. The TDMA approach is often more efficient, as highlighted in [22]: it reduces obviously the collisions. However, a collision-free scheduling is practically complicated to achieve.

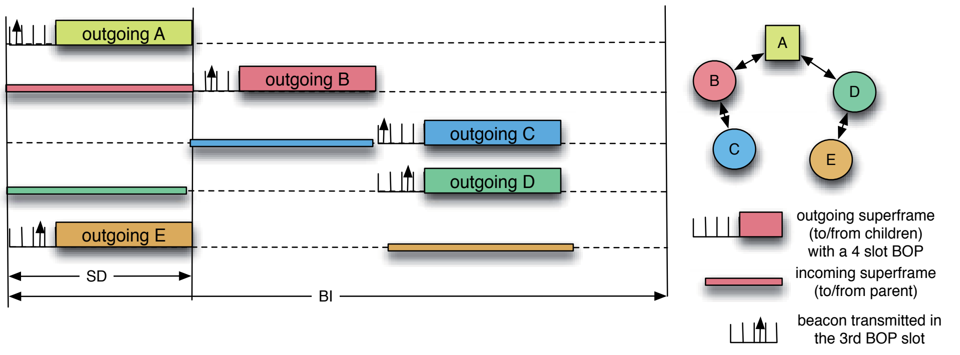


Figure 9. Beacon-Only-Period and superframe scheduling combination

3.4. Combining both techniques

Since we deal with constrained nodes, we aim at minimizing the energy consumption. Thus, we should use shortest paths routing to reduce the number of transmissions. This metric permits to avoid oscillations when creating the DAG while avoiding any loop. We chose to preserve a route length stretch factor of 1: the routes length through the DAG has to be the same as in the original graph. We will adopt the following notation to make a distinction between the different slots:

backoff slot : the units of time used by CSMA-CA algorithm;

GTS slot : the active part of the superframe is divided into slots (which are different from the backoff slots) as described in the standard. The first slots of the active parts are not used, while the last slots correspond to the GTS slots;

BOP slot : the Beacon-Only-Period is divided into slots, containing at most one `beacon`;

superframe slot : a slot containing the whole active part of a superframe, as explained previously.

When the notation is not ambiguous, we will use the term *slot* alone. We integrated both the **Beacon-Only Period** and the **superframe scheduling** techniques as depicted in Figure 9. This organization has the following assets:

1. we avoid wasting bandwidth for coordinators without children;
2. the scheduling permits to maintain several parents, creating redundant paths toward the sink;
3. we reduce energy consumption by maximizing the number of nodes without children.

3.4.1. *Neighborhood discovery*

To implement a distributed scheduling solution, a coordinator should know the list of slots used by interfering coordinators. We consider two coordinators interfere with each other if their superframe collide. For instance, a child of the first coordinator may interfere with the child of the other one. Since determining the exact set of interfering coordinators is complicated to obtain, we should just estimate it. IEEE 802.15.4 [17] proposed to exchange this information along the cluster-tree although the radio and cluster-topologies may very different. We consider rather the 2-radio neighborhood constitutes a better approximation, collisions among 3 or 4-neighbors being solved separately as exceptions.

A coordinator has finally to piggyback in its `beacons` its depth in the cluster-tree (to break conflicts), a leaf flag (i.e. has the coordinator at least one child), the current BOP and superframe slots and a list of 1-neighbors with their short address and their BOP and superframe slots.

To have the radio and not the cluster-tree topology, a coordinator always wakes-up at the beginning of the slots used by its 1-neighbors. It will then go sleeping immediately after having received the corresponding `beacon`. To track changes, a coordinator must periodically stay awake to receive all the `beacons` and update its 2-neighborhood table.

3.4.2. BOP scheduling

The Beacon-Only-Period aims just at solving collisions between coordinators sharing the same superframe. Thus, a new coordinator just picks a random BOP slot not used by any 2-neighbor using the same superframe. We use consequently the approach proposed in [22].

3.4.3. Superframe scheduling

After having collected the slots used by the superframes of its 2-neighboring coordinators, a new coordinator *C* chooses a superframe slot as following:

1. *C* sorts the superframe slots according to the number of 2-neighbors using them;
2. *C* removes the slot of its parent: it cannot maintain a superframe simultaneously without a severe impact on performance;
3. *C* picks randomly one of the least loaded superframe slots.

Obviously, if a superframe slot is not used at all in the 2-neighborhood, it will pick it. However, collisions may even appear because of e.g. inconsistent simultaneous decisions. We can make the distinction between the following cases:

- two coordinators are interfering. In particular, at least one coordinator detects a carrier sense when the other transmits a packet. If they use a different BOP slot, one coordinator will detect an interfering `beacon`: it reapplies the precedent rules forbidding the current slot.

  If they use the same BOP slot, they cannot detect a carrier sense. However, children will not be able to finalize the association since they did not receive the `beacon`. Thus, a coordinator without child (no `association-request` was received) just chooses another superframe slot forbidding the current slot. This change is safe since it does not have any follower.
- an unassociated node suffers from the hidden terminal problem, making a pair of coordinators colliding. With a high probability, this node has another neighboring coordinator with which it can associate. If no candidate exists, the unassociated node will *simulate* a `beacon` transmission to create a collision in the interfering coordinators: at least one of them will choose another superframe slot, solving the collision problem.

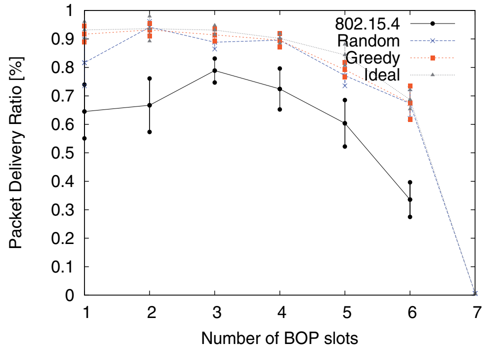


Figure 10. Tuning the Beacon-Only-Period

A coordinator decides to change its slot at the beginning of the active part of its superframe. If the slot has changed, it broadcasts its `beacon` as usually with the new superframe slot and goes sleeping. The possible receivers will be able to detect the source has changed its superframe slot and will update their neighborhood table. They will then *follow* this coordinator (i.e. listen to its new slot to receive its `beacons`).

3.5. Performance evaluation

We used the WSNet simulator [23], with the beacon-enabled IEEE 802.15.4 module [24]. We compared here different scheduling algorithms: original IEEE 802.15.4 (constant `StartTime`), random scheduling (any slot different except the slot of my parent), greedy (the algorithm presented here), ideal (centralized coloring).

We first illustrated the impact of the BOP duration (Figure 10). With only a few BOP slots, collisions between coordinators are more frequent, with a negative impact on the packet delivery ratio. With no surprise, the original IEEE 802.15.4 algorithm performs badly, whatever the BOP duration is. Finally, we can remark having many BOP slots reduces the collision probability but consumes bandwidth: almost all the active part of the superframe is dedicated to `beacons`. We also measured the collision ratio, i.e. ratio of coordinators which have at least one interfering coordinator in the same superframe. We can verify IEEE 802.15.4 presents the largest collision ratio. The greedy scheduling limits collisions, even with low BO values: even when the number of superframe slots is limited, the collision probability remains acceptable. Our greedy solution performs quite close to the ideal scheduling algorithm, which uses centralized a priori knowledge.

4. Optimizing capacity & energy consumption in IEEE 802.15.4 cluster-trees

4.1. Problem statement

Because of the organization into superframes, we have a kind of distributed TDMA scheduling. In particular, all the nodes must transmit their packets to their coordinator (parent) during the active part of the superframe. Since all the active parts are contained into slots of the same size (i.e. Superframe Duration), we may encounter bottlenecks if the cluster-tree is not well balanced. Indeed, too many children in a superframe means all these nodes must share the same radio bandwidth. Moreover, IEEE 802.15.4 is not particularly robust and the number of collisions quickly increases when many nodes participate to the

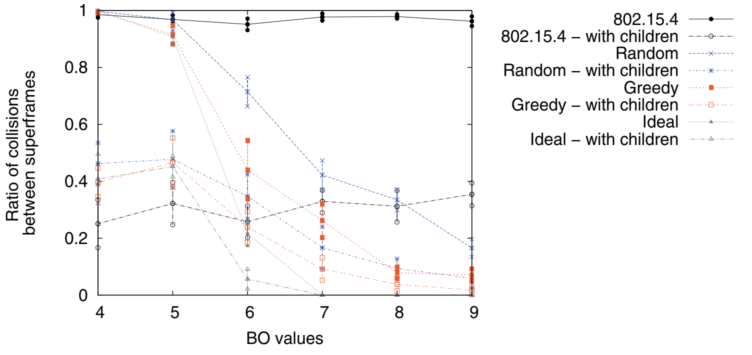


Figure 11. Collisions between superframes

same superframes, leading to collisions, wasting bandwidth. We propose here to formulate this problem as a Mixed Integer Linear Problem (MILP). We will translate all the cluster-tree relations, bandwidth sharing, and IEEE 802.15.4 mechanisms into linear constraints. If we define a linear objective, we face to a classical optimization problem. Our MILP formulation uses the radio and interference topology as an input. Thus, it provides an upper-bound, i.e. what we would obtain with an optimal centralized solution. In particular, we highlighted there exists a large gap to fill: we must still propose a distributed algorithm constructing an efficient cluster-tree.

The standard does not specify which parent to choose when creating the cluster-tree: it just specifies how to associate with. Surprisingly, only few articles investigated the impact of this choice on the cluster-tree topology. Cuomo et al. studied in [8] the average number of children and the height of the tree while Claudios et al. investigated in [25] the dynamic properties of the structure.We propose to investigate how to define formally the cluster-tree structure to create a linear optimization problem.

4.2. Model & assumptions

We expose here all our notations and the models we use : in a heterogeneous IEEE 802.15.4 network, we model the network with a symmetric graph $G = (V, E)$, V being the set of devices and E the set of available links between them. Unidirectional links cannot be exploited in IEEE 802.15.4 and are simply dismissed from the topology. We denote by $N(u)$ the neighborhood of u in G. We associate to G a classical link-interference (or conflict) graph G_c, where each vertex is a radio link, and they are neighbor in G_c if they interfere with each other.

4.2.1. Bandwidth and collisions

As the proposed model must be clearly independent from the PHY layer, we consider the radio bandwidth is equal to 1 unit, and is distributed between interfering transmitters. Collisions may decrease available bandwidth. Even if we assume that `beacons` do not collide during the BOP, CSMA-CA slotted mode does not permit to avoid entirely collisions. This is particularly true when the medium has to forward a large number of packets from different transmitters. A simple and efficient model has been proposed in [26]. Considering collisions are more frequent when the number of transmitters increases, authors demonstrated the percentage of radio bandwidth wasted by collisions is roughly proportional to the number of nodes. Interpolating these results, we estimate the bandwidth wasted due to collisions BW_{coll} follows Equation 3 with a coefficient of determination of 0.97:

$$BW_{coll} = nb\_stations * 0.018675 + 0.0104 \quad (3)$$

where $nb\_stations$ is the number of devices in contention. Note that a superframe should contain at most 53 stations, otherwise all available bandwidth may be wasted by collisions.

4.2.2. Superframes

An efficient way to model the schedule of coordinator's superframes would consist in dividing time into n_{slot} superframes slots. Each active part of a superframe is included in exactly one superframe slot numbered from 0 to $n_{slot} - 1$. Assuming all the nodes have the same beacon period BI and superframe duration SD values (e.g. BO and SO are imposed by the PAN), we have:

$$n_{slot} = \frac{BI}{SD} = 2^{BO-SO}. \quad (4)$$

IEEE 802.15.4 specifies the superframe of one node and its parent must be interspaced by `StartTime`. If we consider `StartTime` is a constant, we should choose `StartTime`=SD. In other words, if a node uses the superframe slot k to transmit its superframe, its child will use the superframe slot $k + 1$. Moreover, we assume that the PAN coordinator will use slot 1. Thus the depth of a node in the cluster-tree gives deterministically its superframe slot. Since the number of timeslots is bounded, we have a kind of modulo: a node with a large depth in the cluster-tree will re-use the superframe slot of the PAN coordinator. If we relax this constraint and and we consider `StartTime` is not anymore a constant (as in Section 3), we have rather to schedule the superframes to limit interference.

4.3. MILP formulation

We present a MILP formulation for the construction of a cluster-tree topology. Input data can be summarized as a collection of devices and the set of all available links represented with a graph $G = (V, E)$ and the corresponding link-interference graph G_c. A $milp$ formulation consists in defining a problem using a set of unaffected variables, a set of linear constraints between them, and an objective function. Then a $milpsolver$ produces a valid variables affectation respecting all the constraint while maximizing / minimizing the objective function. The domain definition and the type of each variable (continuous or integer) must be given as an input to the solver.

4.3.1. Variables

The following part describes all the variables used in our formulation, and their meaning in the resulting affectation. For clarity we have regrouped variables according to their use.

Role definition: we define for each node $u \in V$ a binary variable $\mathbf{root(u)} \in \{0, 1\}$ such that $root(u) = 1$ if and only if u is the network PAN coordinator, and $root(u) = 0$ in all other cases. Any FFD may become coordinator as soon as it is associated to the cluster-tree. Inversely, a RFD is only a leaf of the tree, and cannot accept children. We model this different by using variables $\mathbf{coord(u)} \in \{0, 1\}$, for each node $u \in V$. $coord(u) = 1$ iif u is a FFD.

Superframe Scheduling: in order to assign each coordinator a superframe slot, we introduce integer variables $\mathbf{slot(u)} \in \{0, n_{slot} - 1\}$ for each node $u \in V$ such that $slot(u) = k$ if and only if node u uses the superframe slot number k. When defining the constraints, we need also the additional variables $\mathbf{superframe(u,k)} \in \{0,1\}$ for each node $u \in V$ and each integer $k \in [0, n_{slot}-1]$. $superframe(u,k) = 1$ iif coordinator u is assigned to superframe slot k, or $superframe(u,k) = 0$ otherwise. There is a natural equivalence between $slot(u)$ and $superframe(u,k)$. Indeed, $slot(u) = k$ iif $superframe(u,k) = 1$. We also note that nodes acting as end-devices won't be assigned a superframe slot, thus we state $superframe(u,k) = 0$ for each RFD and each $k \in [0, n_{slot}-1]$ by reducing the definition domain for concerned variables.

Link activity: our main objective consists in determining the subset of links from E restricted to the cluster-tree topology. Only these links may be used to forward the traffic. For each link $(u,v) \in E$ we use a binary variable $\mathbf{child(u,v)} \in \{0,1\}$ such that $child(u,v) = 1$ iif u is a child of node v, and $child(u,v) = 0$ otherwise. A link may be used during the superframe of the corresponding coordinator. We use the binary variables $\mathbf{act(u,v,k)}$ for each radio link $(u,v) \in E$ and each integer $k \in [0, n_{slot}-1]$. $act(u,v,k) = 1$ iif the link (u,v) is active during slot k, i.e. v is a coordinator using timeslot k and u is a child of v. Else, we have $act(u,v,k) = 0$.

Data flow: We define continuous variables $\mathbf{f(u)} \in [0,1]$ as the quantity of data generated by the node u to the PAN coordinator. Then, for each $link(u,v) \in E$ and each integer $k \in [0, n_{slot}-1]$, we note $\mathbf{f(u,v,k)} \in [0,1]$ the traffic from node u to node v using superframe slot k. We also use the continuous variables $\mathbf{f(u,v)} \in [0,1]$ for each link $(u,v) \in E$. $f(u,v)$ is the normalized quantity of bandwidth consumed by the radio link from u to v. It represents the fraction of time during which u transmits its packets to v.

4.3.2. Constraints

We now translate the IEEE 802.15.4 structure and mechanisms into linear constraints.

Tree structure: any device excepted the PAN coordinator has to be associated to exactly one parent node. As $child(u,v)$ are binary integer variables, constraint 5 guarantees the parent unicity. Besides, the PAN coordinator has no parent since it is the root of the cluster-tree (constraint 6). We guarantee the cluster-tree is connected by stating $f(u) > \epsilon$ for each $u \in V - \{PC\}$.

$$\forall u \in V - \{PC\}, \quad \sum_{v \in N(u)} child(u,v) = 1 \tag{5}$$

$$\sum_{v \in N(PC)} child(PC,v) = 0 \tag{6}$$

A node with at least one child must be a coordinator (constraint 7). Reciprocally, a node without child does not maintain a superframe (constraint 8).

$$\forall u \in V, \forall v \in N(u) \quad child(v,u) \leq coord(u) \tag{7}$$

$$\forall u \in V, \quad coord(u) \leq \sum_{v \in N(u)} child(v,u) \tag{8}$$

Superframe scheduling: the following constraint imposes that at most one superframe slot is assigned to any coordinator, and that no slot is assigned to a leaf node:

$$\forall u \in V, \sum_{k \in [0, n_{slot}-1]} superframe(u,k) \leq coord(u) \tag{9}$$

We can also establish a link between variables superframe(u,k) and slot(u) with equation 10:

$$\forall u \in V, \sum_{k \in [0, n_{slot}-1]} superframe(u,k) = slot(u) \quad (10)$$

Each coordinator must be assigned a slot immediately consecutive to it parent, excepted if the parent uses the largest slot number. Else, we have a modulo and the slot number of the child will be 0. We propose to split this problem into 3 constraints:

1. if u is a child of v, the difference between both slots cannot be superior to 2:

$$\forall (u,v) \in E, slot(u) < slot(v) + 2 + (1 - child(u,v)) \times n_{slot} \quad (11)$$

   The part $(1 - child(u,v)) \times n_{slot}$ simply inhibits this constraint when u is not the child of v.

2. in the same way, if u is a child of v, the slot number of u must be at least equal to the slot number of v plus 1, except if u is a leaf (the second line inhibits the constraint in such situation), or if the maximum slot number is assigned to v (third line). Fourth line inhibits this constraint when u is not the child of v.

$$\begin{aligned}\forall (u,v) \in E, slot(u) \geq\ & slot(v) + 1 \\ & - (1 - coord(u)) \times n_{slot} \\ & - superframe(v, n_{slot} - 1) \times n_{slot} \\ & - (1 - child(u,v)) \times n_{slot}\end{aligned} \quad (12)$$

3. Finally, if a node v uses the largest slot ($n_{slot} - 1$), then its child u uses the slot 0:

$$\begin{aligned}\forall (u,v) \in E, \quad slot(u) \leq\ & 2(n_{slot} - 1) - (n_{slot} - 1) \times child(u,v) \\ & - (n_{slot} - 1) \times superframe(v, n_{slot} - 1)\end{aligned} \quad (13)$$

Bandwidth sharing: a node u use the superframe slot of it parent v to exchange packets. Thus, the link (u,v) has be active during the slot k:

$$\forall k \in [0..n_{slot} - 1], \forall (u,v) \in E, f(u,v,k) \leq superframe(v,k) \quad (14)$$

$$\forall (u,v) \in E, \forall k \in [0..n_{slot} - 1], f(u,v,k) \leq act(u,v,k) \quad (15)$$

Then the overall bandwidth from u to v is given as follows :

$$\forall (u,v) \in E, \quad \sum_k f(u,v,k) = f(u,v) \quad (16)$$

Since we consider the superframe slots have the same size, the bandwidth dedicated to one superframe is the radio bandwidth (1 unit) divided by the number of slots. Besides, bandwidth is shared among interfering nodes for their data packets and the corresponding collisions. If we consider the conflict graph, interfering links are neighbors. Thus, we have just to reference all the cliques (interfering links) and verify the bandwidth sharing constraint holds for each clique independently:

$$\forall k \in [0..n_{slot} - 1], \forall c \in \mathcal{C}, \sum_{(u,v) \in c} f(u,v,k) \leq \frac{1}{|n_{slot}|} \left(1 - 0.018675 \sum act(u,v,k) - 0.0104\right) \quad (17)$$

where $\mathcal{C}$ is the set of all the cliques in the interference-link graph G_c.

Flow conservation: Only links between children and parents are available for data flow transmission :

$$\forall (u,v) \in E, \quad f(u,v) \leq child(u,v) \tag{18}$$

Beside, we use the classical flow conservation constraints. Each node sends to its parent both the data packets it generated toward the PAN coordinator and the data packets forwarded from its children:

$$\forall u \in V - \{PC\}, \sum_{p \in N(u)} f(u,p) = \sum_{c \in N(u)} f(c,u) + f(u) \tag{19}$$

Since the PAN coordinator is the single destination (converge case), the following constraint holds:

$$\sum_{u \in N(PC)} f(u,PC) = \sum_{u \in V} f(u) \tag{20}$$

4.3.3. Objective

Our MILP formulation can be associated to any objective defined with linear function. We may maximize the network fair capacity:

$$Obj: \quad max\,(min\,(f(u))) \tag{21}$$

We may also maximize the global network throughput, without guaranteeing fairness, or aim at maximizing the network lifetime by defining linear constraints for the energy consumption and then minimize the maximum energy consumed by each node.

4.4. Numerical results

We illustrate here the performances of the greedy algorithm (a node chooses the first parent to associate with) compared to the upper bound obtained through the MILP formulation (denoted as *optimal* in the graphs). We distribute randomly the nodes in a disk. We consider here 4 timeslots for the superframe scheduling, and use cplex v.12 to solve the MILP formulation. We also plot the 95% confidence intervals. We measure the aggregated network capacity while maintaining the density constant (8 neighbors)(Figure 12). Clearly, associating to the first available parent is sub-optimal: bottlenecks quickly appear, degrading the network capacity. On the contrary, the optimal solution leads to a constant global throughput whatever the number of nodes is (i.e. collisions are very seldom). While this formulation uses the a priori knowledge of the radio and interference topology, we highlighted here the fact that existing cluster-tree construction algorithms are not efficient to optimize the network throughput and the energy consumption.

5. Self-configuring IEEE 802.15.4 version

5.1. Problem statement

IEEE 802.15.4 uses a backoff mechanism to solve contention and retransmissions to cope with the half-duplex and unreliability properties to a radio transmitter. It must cope with any topology while maximizing the throughput and minimizing the energy consumption. However, parameter values greatly impact the performances of the protocol (as highlighted in simulations we will include here). Consequently, we must propose a self-configuration

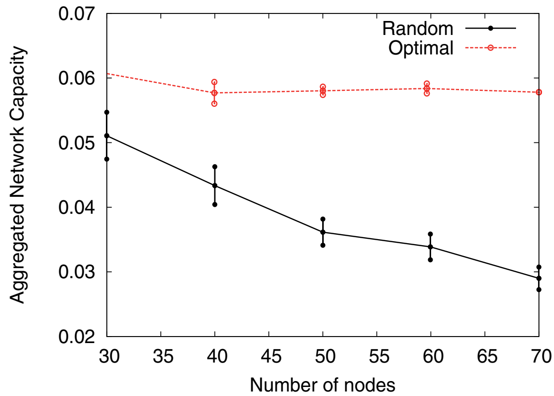


Figure 12. Network capacity vs. nb. of nodes (avg. degree of 8)

method to find always the best parameters values to operate always in optimal conditions. We will present here how we can obtain such property. To optimize the throughput and to reduce the energy consumption, we must reduce the collisions since retransmissions mean bandwidth and energy wastage, and also reduce the backoff times since we waste bandwidth. Obviously, the optimal solution will find the tradeoff between these two antagonist objectives. An analytical model has been presented in [27] to find the optimal parameters values in IEEE 802.15.4. They propose to solve an optimization problem concerning the delay, power-consumption and reliability. The nodes must estimate continuously the busy channel probability and the channel access probability. While IEEE 802.15.4 suggests default values for the parameters (e.g. `BE`, `NB`), they do not lead to optimal performances. Thus, a method has been presented in [18] to maintain a targeted packet delivery radio (PDR). A too large PDR means we waste bandwidth because of large backoff values while a small PDR would not be acceptable for some applications. We consider maximizing the throughput is more efficient: we do not need to drop packets when the traffic is light.

5.2. Defining good parameter values

We propose here to evaluate the impact of each parameter individually.

Contention Window (`CW`): `CW` represents the number of times a node has to trigger a CCA before transmitting its frame. CW should be equal to 2: a node will avoid a collision when the medium is free and the source is waiting for the `acknowledgement`. Thus, the optimal `CW` is its default value.

`macMaxFrameRetries`: `macMaxFrameRetries` denotes the number of times a frame is retransmitted when the `ack` is not received. Anastasi et al. [18] highlighted the fact that packets are very seldom dropped because of too many retransmissions ($<$ 2% when `macMaxFrameRetries`=2).

To optimize the throughput, we may adopt an adaptive approach, depending on the traffic load. When the queue contains many frames to transmit, it should use a small `macMaxFrameRetries`: even if a packet is dropped, others have to be transmitted. On the contrary, a small queue length means we may authorize more retransmissions to *fill* the

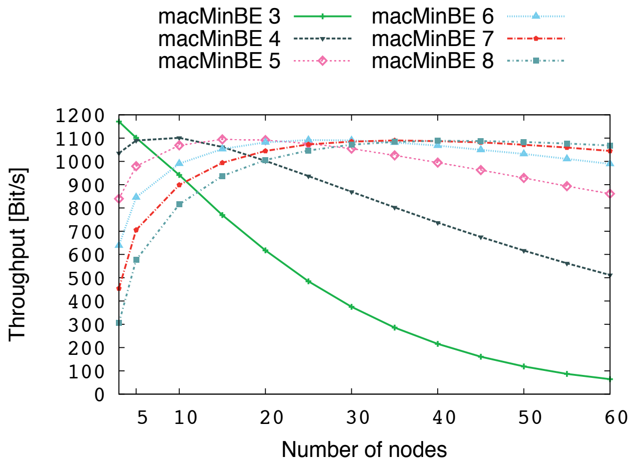


Figure 13. The throughput in saturating mode for IEEE 802.15.4

medium. The standard defining a maximum number of retransmissions of 4, the value may be obtained as following:

$$macMaxFrameRetries = (q_{max} - q(t)) * 4 \tag{22}$$

with q_{max} (resp. $q(t)$) the maximum (resp. current) queue length.

Backoff Exponent (BE): The backoff exponent (BE) defines the range of the random value for the backoff time ($backoff \in [0..2^{BE} - 1]$). Obviously, the backoff exponent has an impact on the throughput. However, its optimal value depends on the number of simultaneous transmitters (cf. Figure 13), which is practically complicated to estimate. We propose in the next section an adaptive approach to always operate close to the optimal BE value.

`macMaxCSMABackoffs`: `macMaxCSMABackoffs` denotes the maximum number of times a node can choose a new backoff value when it senses the medium busy after a CCA. Figure 14) illustrates the impact of `macMaxCSMABackoffs` on the throughput for different traffic loads. A large `macMaxCSMABackoff` value increases the throughput. Indeed, a CCA does not *consume* bandwidth: dropping a packet after a busy CCA is quite aggressive. Obviously, `macMaxCSMABackoff` does increase the throughput only in unsaturated mode: we have to *fill* the whole superframe. We propose the following simple method. When the node has still packets to transmit although the active part has elapsed, it decreases the `macMaxCSMABackoff` value. Inversely, a node which has an empty buffer before the end increases the `macMaxCSMABackoff` value.

5.3. Self-adaptation mechanism for the Backoff Exponent value

To increase the networks lifetime, the nodes must switch their radio off most of the time. The bandwidth is consequently reduced in the same proportion. Besides, all the nodes which participate to a superframe wake-up synchronously and try to send the frames they have buffered since the last beacon of their parent. Thus, even a low traffic may create an high collision rate at the beginning of the active part of the superframe. We must take care of efficiently sharing the bandwidth and reducing the collision rate. We proposed in the previous

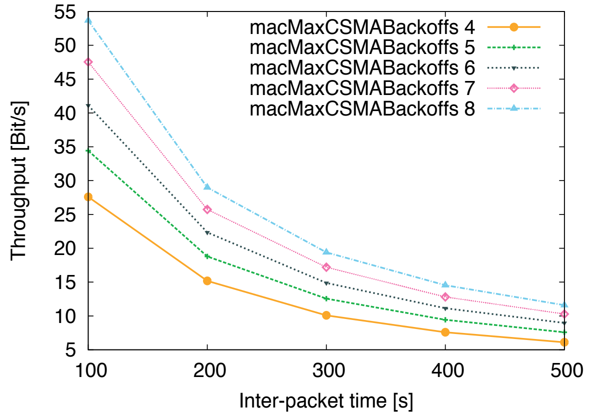


Figure 14. Impact of macMaxCSMABackoff on the throughput

section methods to find the optimal values for most of the parameters. We focus here on the Backoff Exponent, which is practically the most complicated to adjust dynamically.

Our objective consists in limiting bandwidth wastage. A too long backoff leads to medium under-utilization: the nodes wait a too long time, wasting bandwidth. Inversely, a collision *consumes* bandwidth. However, to estimate the sum of bandwidth dedicated to collisions and backoffs is a complicated task. Instead of minimizing bandwidth wastage, we propose rather to maximize the throughput, the antagonist metric. Indeed, we have highlighted the best throughput is stable in Figure 13, i.e. whatever the parameters values, we may obtain at most a throughput of 1,100 b/s. Thus, we *just* have to find the BE value which offers this throughput.

5.3.1. Dynamic exploration

We propose a dynamic exploration algorithm to find the most appropriate BE value. By appropriately testing some *good* BE values, we try to locally optimize the throughput. Indeed, we can remark for a fixed number of nodes, the throughput is monotonically increasing or decreasing when the BE value increases/decreases. We test continuously different values, oscillating around the optimal throughput. Besides, we can remark in Figure 13 the second best BE value leads to a throughput very close to the optimal throughput. Thus, oscillating between both values wastes a limited amount of bandwidth. While the throughput is increasing, we keep on exploring the same direction (either increasing or decreasing the BE value). As soon as the throughput is sub-optimal, we inverse the direction.

5.3.2. Stabilization

A coordinator reaching the optimal throughput should maintain the same BE. Since we have to cope with statistical variations, we consider the throughput has not changed if its difference with the previously measured throughput is inferior to a threshold (Δ= 1 packet/Superframe Duration). To avoid freezing the self-adaptation, a node MUST initiate a new exploration, even if the throughput does not change for a long time. When BE stays unchanged for k successive superframes (3 in our simulations), the coordinator MUST explore another BE value in the same direction as the previous exploration.

Algorithm 1: Updating the BE value

```
1  /* Initialization                                                    */
2  if isCoordinator then
3      myMacMinBE ←3 ;
4      direction ←+1 ;
5      oldTroughput ←0;
6  /* Children read the macMinBE directly in the beacons                */
7  if ! isCoordinator then
8      receiveBeacon;
9      macMinBE ←readValueFromBeacon;
10     return ;
11 /* the coordinator must continuously update its BE                   */
12 if isCoordinator then
13     changeBE = BEUnchangedDuring (k * SD) ;
14     /* the throughput is varying                                      */
15     changeBE |= (newThroughput < oldTroughput- Δ) ;
16     changeBE |= (newThroughput > oldTroughput + Δ) ;
17     if changeBE then
18         /* the throughput has decreased: let change the direction     */
19         if newThroughput < oldTroughput- Δ then
20             direction ←( - direction);
21         myMacMinBE ←myMacMinBE + direction;
22     /* the coordinator sends the BE in beacons                        */
23     sendBeacon(myMacMinBE);
24     oldTroughput ←newThroughput;
```

5.3.3. Consistency

Nodes participating to the same superframe should not use different BE values. Else, IEEE 802.15.4 would become unfair: nodes with a lower BE would monopolize medium access. We propose consequently the BE is announced in the beacons from the coordinator.

The final algorithm is described in Algorithm 1. We use default values on the beginning (lines 2-5). Children extract the BE from beacons and stop any computation (lines 6-10). A coordinator has to continuously update its BE: when the throughput is decreasing, the node changes its direction for the exploration (lines 19-21). When the throughput has changed significantly (more than Δ) or stayed unchanged for a too long time, the BE is increased/decreased, depending on the direction (line 22). Finally, the coordinator saves the current throughput value and piggybacks the new BE in its beacons.

5.4. Performance evaluation

We have evaluated Algorithm 1 to find dynamically the best BE value with the original version of IEEE 802.15.4 with a static BE value. We simulate the protocol with the same parameters as in section 3.5. Figure 15 shows the performance of our protocol (SAV 802.15.4) in saturated mode (i.e. a node has always packets to transmit). We observe that our protocol provides almost optimal throughput, whatever the number of nodes. We also verified our self-adaptation algorithm reacts well to multihop topologies (Figure 16). We can verify the self-adaptation algorithm succeeds to oscillate around the optimal BE and maximizes the throughput.

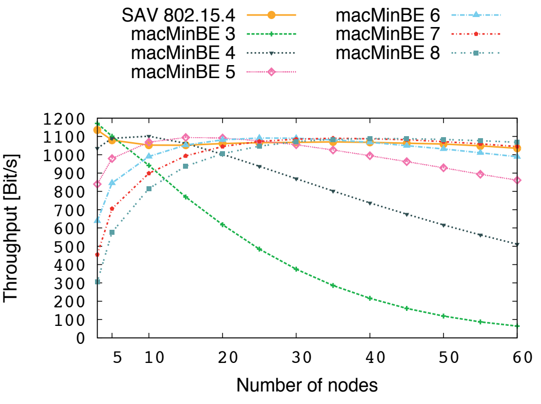


Figure 15. The throughput in saturating mode for IEEE 802.15.4

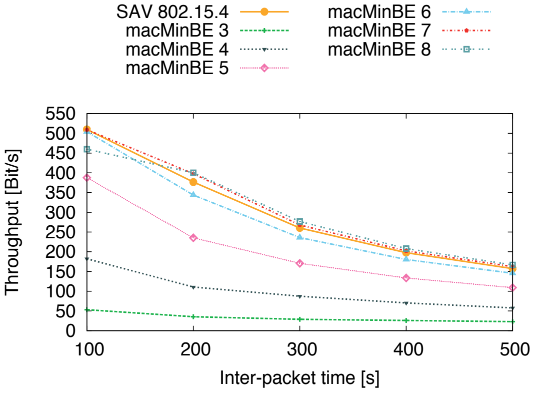


Figure 16. The throughput in non saturating mode of multihop IEEE 802.15.4

6. Open challenges

6.1. Topology formation

In section 4, we have presented a method to evaluate the capacity and energy consumption inherent to any cluster-tree. By providing a tight upper-bound, we can estimate the gap to fill before obtaining an optimal cluster-tree concerning e.g. energy, bandwidth. However, how to reach this objective in a distributed way is still an open problem. Yu *et al.* [28] proposed to minimize the number of clusters: aggregation is simplified, and more wireless routers can sleep. However, we could exploit with benefit the low duty-cycle property of IEEE 802.15.4 to save energy. What strategy is the most efficient in these scenarios? We have also to jointly optimize the scheduling and the cluster-tree hierarchy in a convergecast scenario to minimize the end-to-end delay or at least upper-bounding it. To construct a list of interfering nodes, Tseng *et al.* [29] proposed to use a larger transmission power for neighborhood discovery. However, such process will break the other `data` transmissions impacting negatively the performances. Besides, a trade-off between over and under-estimation is practically complicated to achieve. Moreover, many collisions may occur before converging to a legal scheduling.

6.2. Real-time traffic

IEEE 802.15.4 provides Guaranteed-Time-Slots (GTS) for real time traffic: these slots have to be pre-reserved during the Contention Access Period. In the last version of the standard, the number of GTS can be extended to cope with the particular requirements of the scenario. Park *et al.* [30] studied how to minimize energy consumption for real-time traffic in IEEE 802.15.4 networks. A multihop flow must reserve the GTS along all the path toward the PAN coordinator. Besides, a flow may not require one GTS during each superframe: it depends on the Beacon Order (BO) value and on the data rate of the flow. In that case, the coordinators must multiplex different real-time flows in the same GTS: this corresponds to a bin-packing problem, without knowing a priori the next real-time traffic requests. Finally, some bursty sensitive traffic may be generated because of e.g. alarms. These transmissions should be protected while minimizing the end-to-end delay.

6.3. Metric

A node has to choose the best parent in the cluster-tree. Obviously, it has to define what metric is the most accurate to represent the aptitude of a node to serve as parent. This property depends on selfish (offered bandwidth, reliability, delay) and global metrics (energy consumption toward the PAN coordinator, load-balancing). Thus, a node must take a decision after having captured all these criteria.

6.4. Stability

Testbed often highlighted real conditions lead to instabilities. Bezahaf *et al.* [31] demonstrated routes are very unstable in wireless mesh networks with DSDV. In the same way, Silva *et al.* [32] highlighted the signal is also very time-varying in Wireless Sensor Networks. We must absolutely prevent such oscillations in the cluster-tree. A node may re-associate to another parent because it improves the local and global performances of IEEE 802.15.4. However, it should avoid a domino effect, creating a wave of re-associations.

6.5. Experimental validation

IEEE 802.15.4 has already been extensively studied by simulations: slotted IEEE 802.15.4 was implemented for ns2 in [33], for opnet in [34, 35] (with or without GTS respectively). An analytical framework to model the IEEE 802.15.4 behavior is presented in [36]. Pollin *et al.* [26] presented a Markov Chain for both the saturated and unsaturated traffic scenarios. Chen *et al.* [37] focused on the industrial low rate WPAN and provided an omnet++ implementation. However, only a few real implementations exist. For instance, contiki does not support yet beacon-enabled IEEE 802.15.4. TinyOS has two main implementations:

- open-ZB propose an open source implementation of the beacon-enabled mode of IEEE 802.15.4 with GTS. It also includes the Zigbee features;
- tkn154 includes the full slotted IEEE 802.15.4 except a few features like PAN id conflict notification.

IEEE 802.15.4 has been evaluated with a reactive routing protocol (AODV) on topologies of 6 nodes in [38]. The authors measured only the end-to-end delay for the packets of a flow. The beacon-enabled and non-beacon modes were also compared in a star topology in [39]. More complex scenarios should be evaluated to definitively prove the accuracy of IEEE

802.15.4. In particular, is IEEE 802.15.4 scalable? What are its performances with dozens of hundreds of nodes? Similarly, how does a IEEE 802.15.4 network perform in harsh conditions: outdoor and/or high interference environments, high traffic conditions with large densities, co-existence with other WLAN or WPAN, etc.

7. Conclusion

IEEE 802.15.4 is emerging as the standard for Low-Power Wireless Personal Area Networks. We have presented here how we should modify IEEE 802.15.4 to cope with multihop environments: we must be able to forward traffic while limiting energy consumption. Slots should be carefully scheduled in a distributive way to forward traffic without collisions since it wastes energy. Moreover, a node should choose appropriately its parent to balance the load and energy: the cluster-tree should avoid the creation of bottlenecks. Finally, we have presented a very simple algorithm to find dynamically the best parameters values in IEEE 802.15.4. This enhanced version of IEEE 802.15.4, only modifying slightly the standard optimizes greatly the performances, enabling new types of multihop applications.

Acknowledgement

This work was partially supported by the French National Research Agency (ANR) Contract IRIS ANR-11-INFR-016.

Author details

Abdeddaim Nazim
Grenoble Informatics Laboratory, France

Darties Benoît
LE2I - University of Burgundy, France

Theoleyre Fabrice
CNRS - University of Strasbourg, France

8. References

[1] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon. Health monitoring of civil infrastructures using wireless sensor networks. In *international conference on Information processing in sensor networks (IPSN)*, Cambridge, USA, 2007. ACM.
[2] T. Ulrich. Wireless network monitors h2o: System saves resources, increases yield in cabernet vineyard. *Wines and Vines Magazine*, July 2008.
[3] IEEE 802.15.4-2006 standard. `http://www.ieee802.org/15/pub/TG4.html`,2006.
[4] Joseph Polastre, Jason Hill, and David Culler. Versatile low power media access for wireless sensor networks. In *International conference on Embedded networked sensor systems (Sensys)*, pages 95–107. ACM, Baltimore, MD, USA 2004.
[5] A. Bachir, M. Heusse, A. Duda, and K.K. Leung. Preamble sampling MAC protocols with persistent receivers in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 8(3):1091 –1095, March 2009.
[6] Wei Ye, J. Heidemann, and D. Estrin. An energy-efficient mac protocol for wireless sensor networks. In *INFOCOM*, volume 3, pages 1567 – 1576. IEEE, 2002.

[7] W. Ye and J. Heidemann. Ultra-low duty cycle MAC with scheduled channel polling. In *Conference on Embedded Networked Sensor Systems (SenSys)*, pages 321–334, USA, 2006.

[8] F. Cuomo, E. Cipollone, and A. Abbagnale. Performance analysis of IEEE 802.15.4 wireless sensor networks: An insight into the topology formation process. *Computer Networks*, 53(18):3057–3075, December 2009.

[9] Anna Abbagnale, Emanuele Cipollone, and Francesca Cuomo. A case study for evaluating IEEE 802.15.4 wireless sensor network formation with mobile sinks. In *International Conference on Communications (ICC)*, Dresden, Germany, June 2009. IEEE.

[10] Zigbee. `http://www.zigbee.org/`, 2005.

[11] Meng-Shiuan Pan, Chia-Hung Tsai, and Yu-Chee Tseng. The orphan problem in zigbee wireless networks. *IEEE Transactions on Mobile Computing*, 8(11):1573–1584, November 2009.

[12] N. Hadid, A. Guitton, and M. Misson. Exploiting a meeting channel to interconnect 802.15.4-compliant mobile entities: discovery and association phases. In *International Symposium on Computers and Communications (ISCC)*, pages 94–99. IEEE, june 2010.

[13] Andreas Willig, Niels Karowski, and Jan-Hinrich Hauer. Passive discovery of IEEE 802.15.4-based body sensor networks. *Ad Hoc Networks*, 8(7):742–754, September 2010.

[14] N. Karowski, A. C. Viana, and A. Wolisz. Optimized asynchronous multi-channel neighbor discovery. In *INFOCOM*, Shanghai, China, May 2011. IEEE.

[15] T. Winter, P. Thubert, A. Brandt, J. Hui, R. Kelsey, P. Levis, K. Pister, R. Struik, JP. Vasseur, and R. Alexander. RPL: IPv6 routing protocol for low-power and lossy networks. rfc 6550, 2012.

[16] C. Chaudet, D. Dhoutaut, and I.G. Lassous. Performance issues with IEEE 802.11 in ad hoc networking. *IEEE Communications Magazine*, 43(7):110 – 116, july 2005.

[17] IEEE 802.15 WPAN task group 4b (tg4b). `http://www.ieee802.org/15/pub/TG4b.html`, September 2004.

[18] G. Anastasi, M. Conti, and M. Di Francesco. The mac unreliability problem in IEEE 802.15.4 wireless sensor networks. In *International conference on Modeling, analysis and simulation of wireless and mobile systems (MSWiM)*, pages 196–203, Tenerife, Spain, October 2009. ACM.

[19] Jaejoon Cho and Sunshin An. An adaptive beacon scheduling mechanism using power control in cluster-tree wpans. *Wireless Personal Communications*, 50(2):143–160, 2009.

[20] R. Burda and C. Wietfeld. A distributed and autonomous beacon scheduling algorithm for IEEE 802.15.4/ZigBee networks. In *Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS)*, Pisa, Italy, October 2007. IEEE.

[21] P. Selvan Muthukumaran, R. de Paz, R. Špinar, and D. Pesch. Meshmac: Enabling mesh networking over IEEE802.15.4 through distributed beacon scheduling. In *International Conference on Ad Hoc Networks (AdHocNets)*, Niagara Falls, Ontario, Canada, September 2009.

[22] B.C. Villaverde, R. De Paz Alberola, S. Rea, and D. Pesch. Experimental evaluation of beacon scheduling mechanisms for multihop IEEE 802.15.4 wireless sensor networks. In *SENSORCOMM*, pages 226–231. IARIA, July 2010.

[23] G. Chelius, A. Fraboulet, and E. Fleury. Worldsens: development and prototyping tools for application specific wireless sensors networks. In *International Conference on Information Processing in Sensor Networks (IPSN)*, Cambridge, Massachusetts, USA, 2007. ACM.

[24] N. Abddedaim and F. Theoleyre. Implementation of a wsnet module to simulate the IEEE 802.15.4 beacon-enabled mode in multihop topologies. Research Report 00590853, HAL, May 2011.

[25] F.J. Claudios, R. Radeke, D. Marandin, P. Todorova, and S. Tomic. Performance study of reconfiguration algorithms in cluster-tree topologies for wireless sensor networks. In *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, France, 2007.

[26] S. Pollin, M. Ergen, S. Ergen, B. Bougard, L. Der Perre, I. Moerman, A. Bahai, P. Varaiya, and F. Catthoor. Performance analysis of slotted carrier sense IEEE 802.15.4 medium access layer. *IEEE Transactions on Wireless Communications*, 7(9):3359 –3371, September 2008.

[27] Pangun Park, Carlo Fischione, and Karl Henrik Johansson. Adaptive IEEE 802.15.4 protocol for energy efficient, reliable and timely communications. In *International Conference on Information Processing in Sensor Networks (IPSN)*, pages 327–338, Stockholm, Sweden, 2010. ACM/IEEE.

[28] C. W. Yu, C. Chang, and J. Chou. Minimizing the number of clusters in IEEE 802.15.4 wireless sensor networks. In *International Workshop on Wireless & Mobile Networks (WiMoN)*, 2009.

[29] Yu-Chee Tseng and Meng-Shiuan Pan. Quick convergecast in ZigBee/IEEE 802.15.4 tree-based wireless sensor networks. In *International workshop on Mobility management and wireless access (MobiWac)*, pages 60–66, Terromolinos, Spain, 2006. ACM.

[30] Myung-Gon Park, Kang-Wook Kim, and Chang-Gun Lee. Holistic optimization of real-time IEEE 802.15.4/ZigBee networks. In *International Conference on Advanced Information Networking and Applications (AINA)*, pages 443 –450. IEEE, march 2011.

[31] M. Bezahaf, L. Iannone, M. Dias de Amorim, and S. Fdida. Insights into the routing stability of a multi-hop wireless testbed. In *International Conference on Ad Hoc Networks (AdHocNets)*, Victoria, BC, Canada, 2010.

[32] A. R. Silva and M. C. Vuran. Development of a testbed for wireless underground sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 2010:9:1–9:12, April 2010.

[33] Jianliang Zheng and Myung J. Lee. *Sensor Network Operations*, chapter A Comprehensive Performance Study of IEEE 802.15.4, pages 218–237. IEEE Press, 2006.

[34] Anis Koubaa, Mario Alves, and Eduardo Tovar. A comprehensive simulation study of slotted CSMA/CA for IEEE 802.15.4 wireless sensor networks. In *International Workshop on Factory Communication Systems (WFCS)*, Torino, Italy, June 2006. IEEE.

[35] P. Jurcik, A. Koubaa, M. Alves, E. Tovar, and Z. Hanzalek. A simulation model for the IEEE 802.15.4 protocol: Delay/throughput evaluation of the GTS mechanism. In *MASCOTS*, oct. 2007.

[36] Tae Ok Kim, Jin Soo Park, Kyung Jae Kim, and Bong Dae Choi. Analytical model of IEEE 802.15.4 non-beacon mode with download traffic by the piggyback method. In *Network and Parallel Computing (NPC)*, pages 435–444, Dalian, China, 2007.

[37] Feng Chen, Nan Wang, R. German, and F. Dressler. Performance evaluation of IEEE 802.15.4 LR-WPAN for industrial applications. In *Conference on Wireless on Demand Network Systems and Services (WONS)*, pages 89 –96, Garmisch-Partenkirchen, Germany, jan. 2008. IEEE.

[38] C. Gomez, P. Salvatella, O. Alonso, and J. Paradells. Adapting AODV for IEEE 802.15.4 mesh sensor networks: theoretical discussion and performance evaluation in a real environment. In *International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Niagara-Falls, USA, June 2006. IEEE.

[39] C. Suh, Z. Mir, and Y. Ko. Design and implementation of enhanced IEEE 802.15.4 for supporting multimedia service in wireless sensor networks. *Computer Networks*, 52(13), September 2008.

Chapter 6

Integrated Reliability Modelling for Wireless Sensor Networks

Qiang Zhou

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48429

1. Introduction

This chapter deals with the problem of reliability modelling with wireless sensor network (WSN) (Akyildiz et al., 2002), which is rapidly becoming a platform for applications including antiterrorism, smart spaces, numerous military sensing and command and control applications, and entertainment. Inherent in these safety-critical applications is a priority and urgency of the information or messages.

There are significant research results on reliability modelling for distributed systems with wired network. (Lin, 1976) approaches a connectivity-based reliability modelling from the perspective of the networks, which consider the node unit and link unit. (Tripath, 1997) proposes task-based reliability modelling by just considering processor unit, and defines a reliability index for a task, but don't cover the system reliability perspective. (Zhou et al., 2001) approaches the reliability modelling from the perspective of the nodes in which a task involves and uses the reliability matrix with each element as the reliability of a task to evaluate reliability of a distributed system in avionics platform.

However, as far as we know, few attention has been paid by researchers to addressing reliability modelling on WSNs. Especially for safety-critical application, the reliability is influenced mainly and directly by not only the connectivity of the network topology but also the (such as energy-/time-) efficiency of the system. (Feng & Kumar, 2004) researches the connectivity reliability of wireless networks, but don't consider the reliability modelling. (Xing & Shrestha, 2006) considers the problem of reliability modelling and analysis of hierarchical clustered wireless sensor networks (WSN), proposes reliability measures that integrate the conventional connectivity-based network reliability with the sensing coverage measure indicating the quality of service (QoS) of the WSN. Both work above research reliability problem of WSN just from view of connectivity and coverage, don't introduce the efficiency of the system. (Xing, 2006; AboElFotoh et al., 2005) consider the efficiency in reliability modelling for WSN. (Xing, 2006) proposes an integrated modelling on WSN reliability and security. (AboElFotoh et al., 2005) considers the delay-efficiency factor into the reliability modelling, by computing a measure for the reliability and a measure for the message delay between data sources & data sinks in an WSN, respectively. (Silva et al., 2012) proposes a methodology based on an automatic generation of a fault tree to evaluate the reliability and availability of WSNs, when permanent faults occur on network devices. (Johannes et al., 2012) generalizes the expected hop count metric (EHC) into an expected message delay (EMD) that permits arbitrary delay values for both links and devices. Further, it proposes a method based on Augmented Ordered Multivariate Decision Diagram (OMDD-A) that can be used to compute reliability (REL), EHC and EMD for WSN with both device and link failures.

To the best of our knowledge, however, there is no systematical research done to unify energy consume and message delay into reliability modelling for WSN. The work in this chapter differs from the previous work in that it proposes a model of the system and an integrated model of the task which considers energy consume and message delay for the safety-critical application, introduces both the energy factor function and time factor function, and also establishes an integrated reliability model of WSN based on a task. The illustration of modelling suggests that the method studied has a directive influence to both task division and topology selection of WSN system.

The rest of the chapter is organized as follows. The basic node model and network structure of wireless sensor networks are introduced in Section 2. We propose an integrated model of the task which considers energy consume and message delay based on a task, introduce both the energy factor function and time factor function, and also establish an integrated reliability model of WSN based on a task in Section 3 and 4, respectively. In Section 5, we present an illustration of modelling of representative hierarchical cluster topology in WSN. Finally, the chapter is concluded in Section 6.

2. The network structure and node model

2.1. The network structure

WSNs composed of multiple sensor nodes and one *sink node*. The sensor nodes are usually scattered in a *sensor field* as in Fig.1. Each of these scattered sensor nodes has the capabilities to collect data and route data back to the *sink node*. Data are routed back to the sink node by a multi-hop architecture through the sink node as shown in Fig.1.

2.2. Node model

A sensor node is made up of four basic components as shown in Fig. 1: an acquisition unit, a processing unit, a communication unit and a power unit. Acquisition units are usually composed of two subunits: sensors and analog to digital converters (ADCs). The analog signals produced by the sensors based on the observed phenomenon are converted to digital signals by the ADC, and then fed into the processing unit. The processing unit, which is generally associated with a small storage unit, manages the procedures that make the sensor node collaborate with the other nodes to carry out the assigned sensing tasks. A communication unit connects the node to the network. One of the most important components of a sensor node is the power unit. Power units may be supported by a power scavenging unit such as solar cells. There are also other subunits, which are application dependent.

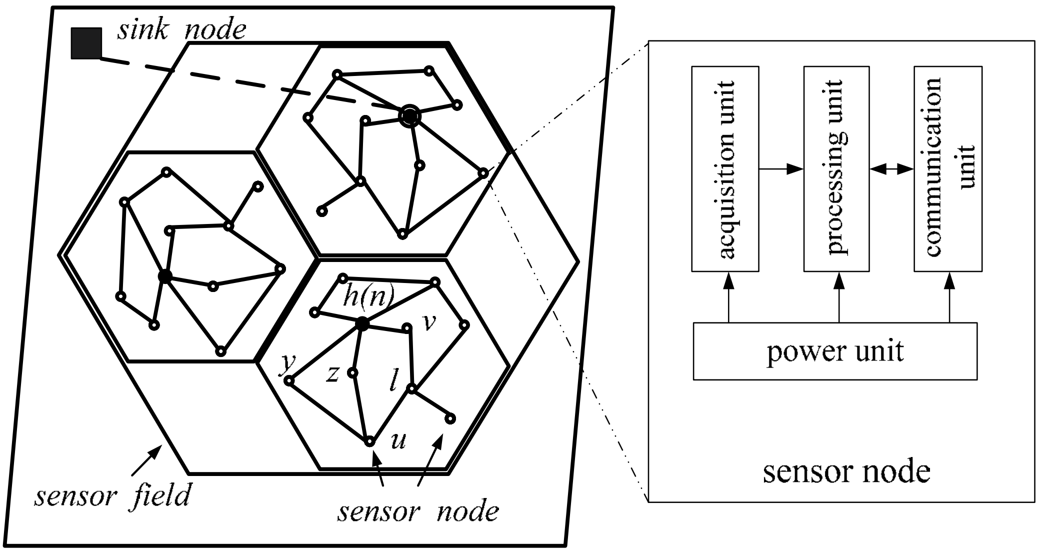


Figure 1. Wireless sensor node model and network structure

Data interchanges between nodes are generally supported by processing units and communication units. Because those acquisition units can only have effect in source nodes, instead of taking effect on data relay, we suppose the acquisition unit works well. Besides, we also suppose the power unit works well.

3. The network task models

WSN can support a serial of data interchanges to satisfy the functional need for some applications. Moreover, the process of every such data interchange can be looked as a task message, which includes message source, message route, message destination and so on. So, the whole WSN can be functionally looked as a task set.

In this Section, we propose an integrated model of the task which considers energy consume and message delay for the safety-critical application, , and introduce both the energy factor function and time factor function.

3.1. Task model

[definition A]: A task in WSN can be characterized as $\tau_{S,D,E,T}$, where

- S : represents a source node;
- D : represents a destination node;
- E : represents the maximum value of all energy consumer to transmit the task message from one node to another in a single-hop way through the route;
- T : represents the amount of time to transmit the task message from the source node to the destination one.

3.2. Route set of a task

[definition B]: when a task $\tau_{S,D,E,T}$ is executed, task message can bypass several node orderly, so the ordered node set composes a route of the task $\tau_{S,D,E,T}$, which characterized as $r_{S,D}$. Where, $r_{S,D}$ just consider the connectivity of the task instead of energy/time constraint. Generally speaking, there maybe more than one route for a task in WSN, so we characterize the number of the route of the task $\tau_{S,D,E,T}$ as $k_{S,D}$.

3.2.1. Energy available route

The wireless sensor network (WSN) system is a kind of system in which the consumer energy of a task is strictly bounded. So, an energy factor function is defined to decide whether the energy is available for a route.

- Energy factor function

Suppose that task $\tau_{S,D,E,T}$ has one route $r^i_{S,D}$, which has the ordered nodes as follows $n^i_1, n^i_2, ..., n^i_j, ..., n^i_h$. Where h represents the number of nodes bypassed by task message. We denote $E_{n^i_j, n^i_{j+1}}$ as the consumer energy for message transmission from node n^i_j to node n^i_{j+1} of route $r^i_{S,D}$, and denote $E_{n^i_j}$ as the available power in node n^i_j, that means $E_{n^i_j}$ is the remaining energy of n^i_j (Sun et al., 2005). Then, the *energy factor function* can be defined as follows:

$$F(\varepsilon_i) = \begin{cases} 1 & E_{n^i_j, n^i_{j+1}} \le E_{n^i_j} \quad (j = 1, 2, ..., h-1) \\ 0 & \text{other case} \end{cases} \tag{1}$$

The expression (1) indicates that route $r^i_{S,D}$ is an *energy available route* only if that the available power $E_{n^i_j}$ in each node is no less than $E_{n^i_j, n^i_{j+1}}$. Specifically, $\tau_{S,D,E,T}$ can be denoted as $\tau_{S,D,0,T}$, if $E = 0$. This means that the energy constrains can be ignored.

3.2.2. Delay available route

The wireless sensor network (WSN) system is also a kind of system in which the message delay of a task is strictly bounded. So, a time factor function is defined to decide whether the delay is available.

- Time factor function

Suppose that task $\tau_{S,D,E,T}$ has one route $r^i_{S,D}$, which is defined as in section 3.2.1. We denote $T_{n^i_j,n^i_{j+1}}$ as the consumer time for message transmission from node n^i_j to node n^i_{j+1} of route $r^i_{S,D}$, and denote t_i as the sum of $T_{n^i_j,n^i_{j+1}}$ for route $r^i_{S,D}$. Then, the *time factor function* can be defined as follows:

$$F(t_i)=\begin{cases}1 & t_i \le T\\ 0 & t_i > T\end{cases} \tag{2}$$

The expression (2) indicates that route $r^i_{S,D}$ is a *delay available route* only if that the transmission delay t_i is no more than message deadline T. Specifically, $\tau_{S,D,E,T}$ can be denoted as $\tau_{S,D,E,\infty}$, if $T=\infty$. This means that the time constrains can be ignored. In another word, $\tau_{S,D,E,\infty}$ will consider the energy constrains, while $\tau_{S,D,0,T}$ will consider the time constrains. Moreover, $\tau_{S,D,E,T}$ can be denoted as $\tau_{S,D,0,\infty}$ if both the time constrains and the energy constrains are ignored.

3.3. Available route set of a task

So, consider both the energy factor and the time factor, if the route $r^i_{S,D}$ can meet both energy constraint and time constraint, it is called an *available route*, denoted as $r^i_{S,D,E,T}$. And then, the *available route set* of $\tau_{S,D,E,T}$ is:

$$R_{S,D,E,T}=\sum_{i=1}^{k_{S,D}} r^i_{S,D,E,T}=\sum_{i=1}^{k_{S,D}} F(\varepsilon_i)\cdot F(t_i)\cdot r^i_{S,D} \tag{3}$$

Where the route $r^i_{S,D,E,T}$ of the task $\tau_{S,D,E,T}$ composed of a processing unit set and a communication unit set, which are denoted as $P^i_{r_{S,D,E,T}}$ and $C^i_{r_{S,D,E,T}}$, respectively.

$$P^i_{r_{S,D,E,T}}=\left\{p^i_{r_{S,D,E,T}} \mid p^i_{r_{S,D,E,T}} \in r^i_{S,D,E,T}\right\} \tag{4}$$

$$C^i_{r_{S,D,E,T}}=\left\{c^i_{r_{S,D,E,T}} \mid c^i_{r_{S,D,E,T}} \in r^i_{S,D,E,T}\right\} \tag{5}$$

Therefore, the processing unit set and the communication unit set of $\tau_{S,D,E,T}$ can be denoted as, respectively

$$P_{T_{S,D,E,T}}=\bigcup_{i=1}^{k_{S,D}} P^i_{r_{S,D,E,T}} \tag{6}$$

$$C_{T_{S,D,E,T}}=\bigcup_{i=1}^{k_{S,D}} C^i_{r_{S,D,E,T}} \tag{7}$$

4. Reliability model

In this Section, we establish an integrated reliability model of WSN based on task.

4.1. Assumptions

Given the WSNs nodes has a number size of *M*.

1. We take the assumption that the occurrence of component failures is independent, components either work or fail.
2. Assume the variable *m* is the number of tasks; therefore, the expression of the system task set is $\Gamma = \left\{ \tau_{S,D,E,T}^{k} \middle| k = 1,2,...,m \right\}$.

4.2. Task reliability

In WSNs, task $\tau_{S,D,E,T}$ can have more than one route. *The reliability of the route* $r_{S,D,E,T}^{i}$ of task $\tau_{S,D,E,T}$ is equivalent to the probability of the processing units $P_{r_{S,D,E,T}}^{i}$ and communication units $C_{r_{S,D,E,T}}^{i}$ working properly. That is:

$$R_{r_{S,D,E,T}^{i}} = Pr\left\{ r_{S,D,E,T}^{i} \right\} = Pr\left(P_{r_{S,D,E,T}}^{i} \right) \cdot Pr\left(C_{r_{S,D,E,T}}^{i} \right) \tag{8}$$

Where, *Pr{•}* denotes the probability of the object' s working properly in above bracket.
Task reliability is equivalent to the probability that there exists at least one path among Task paths, which is

$$R_{\tau_{S,D,E,T}} = Pr\left\{ \bigcup_{i=1}^{k_{S,D}} r_{S,D,E,T}^{i} \right\} \tag{9}$$

According to the formula of probability for incompatible event (Zhang et al., 1997), we have

$$\begin{aligned} R_{\tau_{S,D,E,T}} &= Pr\left\{ \bigcup_{i=1}^{k_{S,D}} r_{S,D,E,T}^{i} \right\} = \sum_{i=1}^{k_{S,D}} Pr\left\{ r_{S,D,E,T}^{i} \right\} - \sum_{i<j=2}^{k_{S,D}} Pr\left\{ r_{S,D,E,T}^{i} \bigcap r_{S,D,E,T}^{j} \right\} \\ &+ \sum_{i<j<h=3}^{k_{S,D}} Pr\left\{ r_{S,D,E,T}^{i} \bigcap r_{S,D,E,T}^{j} \bigcap r_{S,D,E,T}^{h} \right\} + \cdots + (-1)^{k_{S,D}-1} \cdot Pr\left\{ \bigcap_{i=1}^{k_{S,D}} r_{S,D,E,T}^{i} \right\} \end{aligned} \tag{10}$$

5. Examples

In this Section, we present an illustration of modelling of representative hierarchical cluster topology in WSN, as shown in Figure 2.

WSN can offer unprecedented flexibility in the choice of network topology to match the mission requirements, and a large number of network topology architectures have been proposed for WSNs (Tilak et al., 2002; Edgar et al., 2003), and a topology solution that is efficient for one architecture is likely not to be the best for another, as different network architectures exhibit different communication patterns. Therefore, the topology selection and reliability evaluation are important issues for distributed WSN.

Presently, mesh and hierarchical clustered topology have emerged as the choice topologies for sensor networks. To decrease communication traffic and communication frequency, to ensure scalability and fault tolerance, and to manage the large number of sensors, WSN use the clustered hierarchical architecture (Heinzelman et al., 2000; Tubaishat & Madria, 2003).

5.1. Hierarchical clustered topology

Figure 2 shows an example hierarchical clustered structure with nodes organized into different layers (Banerjee & Khuller, 2001; Tubaishat et al., 2003; Kim, 2010). All the sensor nodes in the network are joined at the lowest layer. The cluster heads in layer-0 are arranged into clusters in layer-1 and a cluster head is assigned for each cluster at this layer. The process is repeated for each layer until the highest layer in the architecture is reached. The hierarchical scheme forms a tree structure for routing with the sink node as the root of the tree (Callaway, 2004). Whenever a sensor node needs to send a message to the sink or another sensor node, it sends the message to its cluster head. The message is routed progressively to the immediately higher-level cluster heads, each of which forms a more detailed segment of the route, until it reaches the cluster head that has the routing information about the destination node. The message is then routed progressively to lower-level cluster heads until it reaches the destination node.

Base on the reliability modelling method above, we analyze the reliability of the hierarchical cluster structure of WSNs. The analysis includes two parts, 1) all cluster heads from L_1 to L_n ; 2) the lowest cluster L_n .

Suppose a task $\tau_{i,sink}$, where sink represents the destination node, and i represents the source node which belongs to the lowest cluster L_n . Based on the characteristic of hierarchical cluster structure, the routes of $\tau_{i,sink}$ have the ordered cluster layers as $(L_n \to L_{n-1} \to ... \to L_1 \to sink)$ which is indicated in Table 1. Moreover, the cluster heads in each layer forms a tree structure for routing with the sink node as the root of the tree. Correspondingly, the part of the ordered cluster heads in a route of the task $\tau_{i,sink}$ is always $(h_{n-1},...,h_1,sink)$. By contrary, the part of the lowest cluster in the route may have multiple sub-routes such as $(1,2,3,6,h_n)$, $(1,2,5,8,h_n)$, $(1,4,5,6,h_n)$, and etc. (see Table 1 and Fig.2.)

So, the reliability of task $\tau_{i,sink}$ can be presented as

$$R_{\tau_{i,sink}} = R_{\tau_{h(n-1),sink}} \bigcap R_{\tau_{i,h(n)}} \tag{11}$$

Where n represents the cluster depth; $h(w) = h_w \ (w = 1,...,n)$;

$R_{\tau_{h(n-1),sink}}$ represents the reliability of the part of the ordered cluster heads in the route;

$R_{\tau_{i,h(n)}}$ represents the reliability of the part of the lowest cluster in the route.

Therefore, by (8), we can deduce the following expression

$$R_{\tau_{h(n-1),sink}} = Pr\left(P_{r_{h(n-1),sink}}\right) \cdot Pr\left(C_{r_{h(n-1),sink}}\right) \tag{12}$$

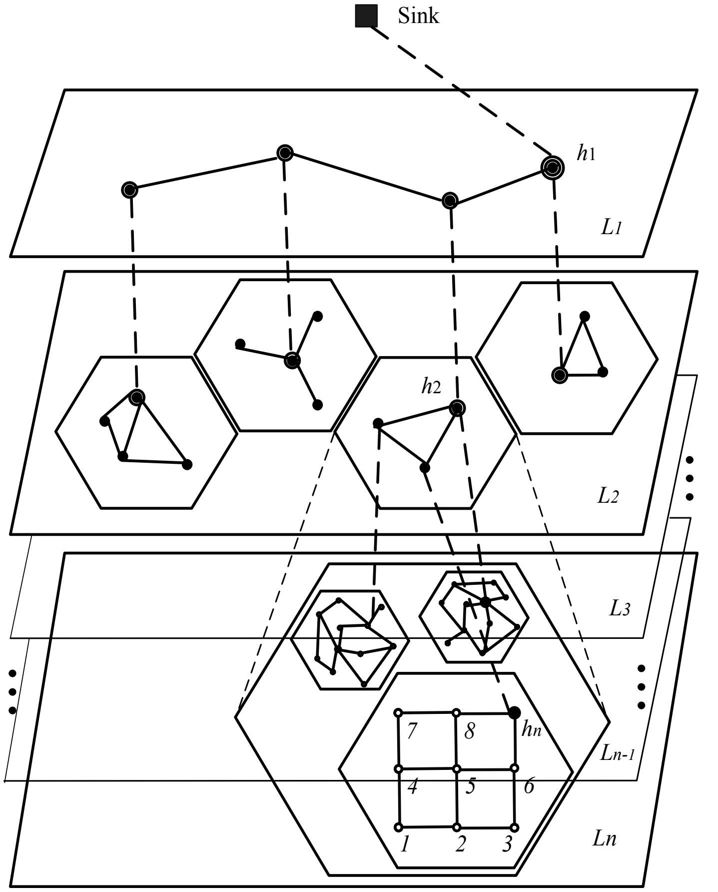


Figure 2. Hierarchical cluster structure of WSNs

No.	Route	Clusters to route
1	1,2,3,6, h_n, h_{n-1},…, h_1,sink	L_n->L_{n-1}->…-> L_1->sink
2	1,2,5,8, h_n, h_{n-1},…, h_1,sink	L_n->L_{n-1}->…-> L_1->sink
3	1,4,5,6, h_n, h_{n-1},…, h_1,sink	L_n->L_{n-1}->…-> L_1->sink
…	Other route	L_n->L_{n-1}->…-> L_1->sink

Table 1. Routes of $\tau_{i,sink}$ (i=1) and its clusters

5.2. Reliability analysis

Example: Suppose a task set $\Gamma = \{\tau_{1,sink,2,16}, \tau_{3,sink,2,16}, \tau_{7,sink,2,16}\}$ in Fig.2, we will analyze the reliability of every task in Γ by using the reliability modelling method proposed in the section above. And the analysis content includes:

- To analyze the routes of every task in Γ in the lowest cluster.
- To analyze the constrained energy'/time' influence on task reliability/system reliability.

Settings and assumptions:

- The cluster depth $n = 5$;
- For the part of the ordered cluster heads in the route, it is supposed that: $E_{h(n),h(n\text{-}1)} = E_{h(n-1),h(n-2)} = \ldots = E_{h(2),h(1)} = E_{h(1),sink} = 2$; $E_{h(n)} = E_{h(n-1)} = \ldots = E_{h(1)} = E_{sink} = 2$; $T_{h(n),h(n\text{-}1)} = \ldots = T_{h(2),h(1)} = T_{h(1),sink} = 2$. Therefore, by (12), we can get the sub-task $\Gamma^{'} = \{\tau_{1,h(n),2,8}, \tau_{3,h(n),2,8}, \tau_{7,h(n),2,8}\}$ corresponding to the lowest cluster.
- For the part of the lowest cluster in the route, as in Fig.3, it is supposed that: $E_1 = E_3 = E_4 = E_5 = E_6 = E_8 = 2$; $E_2 = E_7 = 1$; $E_{2,3} = E_{2,5} = E_{5,6} = E_{7,8} = 1.5$; $E_{1,2} = E_{1,4} = E_{4,7} = E_{5,8} = E_{4,5} = E_{3,6} = E_{8,h(n)} = E_{6,h(n)} = 2$; $T_{7,8} = 1$; $T_{5,8} = T_{5,6} = T_{3,6} = 3$; $T_{1,2} = T_{2,3} = T_{1,4} = T_{2,5} = T_{4,5} = T_{4,7} = T_{6,h(n)} = T_{8,h(n)} = 2$.
- all processing units and communication units have the same failure rate, in order to reflect the influence of the failure rate to the system efficiently, the range of the failure rate is [1e-1,1e-3].

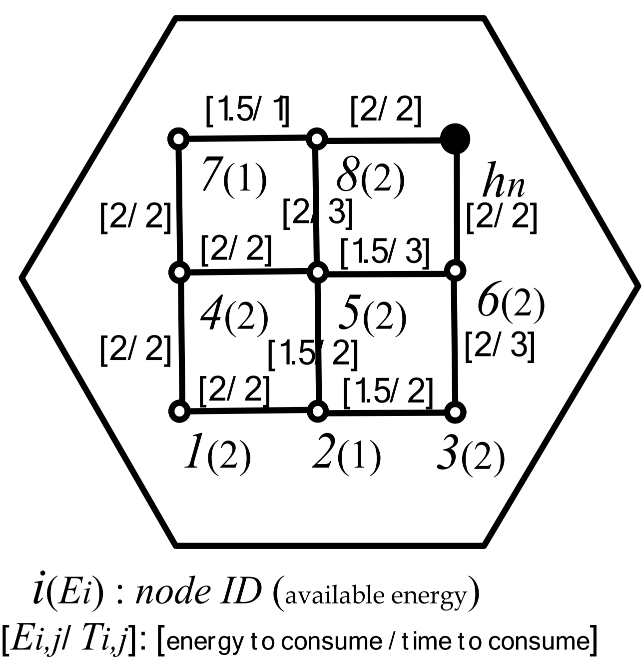


Figure 3. Nodes in the lowest cluster and its energy/time setting

Reliability analysis:

According to both the reliability modeling method and the routing scheme (see Fig.2) in the lowest cluster, we firstly obtain all the routes of each task, then obtain both the processing unit set and the communication unit set of those routes, after that search the energy available route and the time available route of each task, and finally analyze the reliability index of each task.

Table 2 shows that multiple routes exist when considering the connectivity of the task $\tau_{1,h(n),2,8}$; that only route 4 and route 5 can meet energy constraint, so they are *energy available routes*; and that only route 6 can meet time constraint, so it is a delay available route. Further more, there no route can meet both energy constraint and time constraint, so there no available route of the task.

In the same way, table 3 and table 4 show the route and its availability of $\tau_{3,h(n),2,8}$ and $\tau_{7,h(n),2,8}$, respectively.

By (8),(9),and (10), we can deduce the reliability index of these three tasks. Taking $\tau_{3,h(n),2,8}$ as an example, when just considering the *energy available route* of route 1 and route 2, its reliability index can be calculated as

$$\begin{aligned} R_{\tau_{3,h(n),2,\infty}} &= Pr\left\{\bigcup_{i=1}^{2} r^{i}_{3,h(n),2,\infty}\right\} = Pr\left(r^{1}_{3,h(n),2,\infty}\right) + Pr\left(r^{2}_{3,h(n),2,\infty}\right) - Pr\left(r^{1}_{3,h(n),2,\infty} \cap r^{2}_{3,h(n),2,\infty}\right) = \\ &= Pr\left(P^{1}_{r_{3,h(n),2,\infty}}\right)\cdot Pr\left(C^{1}_{r_{3,h(n),2,\infty}}\right) + Pr\left(P^{2}_{r_{3,h(n),2,\infty}}\right)\cdot Pr\left(C^{2}_{r_{3,h(n),2,\infty}}\right) - \\ &- Pr\left(P^{1}_{r_{3,h(n),2,\infty}} \cap P^{2}_{r_{3,h(n),2,\infty}}\right)\cdot Pr\left(C^{1}_{r_{3,h(n),2,\infty}} \cap C^{2}_{r_{3,h(n),2,\infty}}\right) \end{aligned} \tag{13}$$

And when considering the (energy and time) *available route* of route 1, its reliability index can be calculated as

$$R_{\tau_{3,h(n),2,8}} = Pr\left\{r^{1}_{3,h(n),2,8}\right\} = Pr\left(P^{1}_{r_{3,h(n),2,8}} \cap C^{1}_{r_{3,h(n),2,8}}\right) \tag{14}$$

No.	route	Processing unit set	Communication unit set	$f(\varepsilon_i)$	$f(t_i)$	Available route?
1	$1,2,3,6,h_n$	$p_1,p_2,p_3,p_6,p_{h(n)}$	$c_1,c_2,c_3,c_6,c_{h(n)}$	0	0	N
2	$1,2,5,6,h_n$	$p_1,p_2,p_5,p_6,p_{h(n)}$	$c_1,c_2,c_5,c_6,c_{h(n)}$	0	0	N
3	$1,2,5,8,h_n$	$p_1,p_2,p_5,p_8,p_{h(n)}$	$c_1,c_2,c_5,c_8,c_{h(n)}$	0	0	N
4	$1,4,5,8,h_n$	$p_1,p_4,p_5,p_8,p_{h(n)}$	$c_1,c_4,c_5,c_8,c_{h(n)}$	**1**	0	**N**
5	$1,4,5,6,h_n$	$p_1,p_4,p_5,p_6,p_{h(n)}$	$c_1,c_4,c_5,c_6,c_{h(n)}$	**1**	**0**	**N**
6	$1,4,7,8,h_n$	$p_1,p_4,p_7,p_8,p_{h(n)}$	$c_1,c_4,c_7,c_8,c_{h(n)}$	0	**1**	N
…	Other route	…	…	0	0	N

Table 2. Routes of $\tau_{1,h(n),2,8}$ and its processing/communication unit set

No.	route	Processing unit set	Communication unit set	$f(\varepsilon_i)$	$f(t_i)$	Available route?
1	$3,6,h_n$	$p_3,p_6,p_{h(n)}$	$c_3,c_6,c_{h(n)}$	**1**	**1**	**Y**
2	$3,6,5,8,h_n$	$p_3,p_6,p_5,p_8,p_{h(n)}$	$c_3,c_6,c_5,c_8,c_{h(n)}$	**1**	**0**	**N**
3	$3,6,5,4,7,8,h_n$	$p_3,p_6,p_5,p_4,p_7,p_8,p_{h(n)}$	$c_3,c_6,c_5,c_4,$ $c_7,c_8,c_{h(n)}$	0	0	N
4	$3,6,5,2,1,4,7,8,h_n$	$p_3,p_6,p_5,p_2,p_1,p_4,p_7,$ $p_8,p_{h(n)}$	$c_3,c_6,c_5,c_2,c_1,c_4,c_7,$ $c_8,c_{h(n)}$	0	0	N
5	$3,2,5,6,h_n$	$p_3,p_2,p_5,p_6,p_{h(n)}$	$c_3,c_2,c_5,c_6,c_{h(n)}$	0	0	N
6	$3,2,5,8,h_n$	$p_3,p_2,p_5,p_8,p_{h(n)}$	$c_3,c_2,c_5,c_8,c_{h(n)}$	0	0	N
7	$3,2,5,4,7,8,h_n$	$p_3,p_2,p_5,p_4,p_7,p_8,p_{h(n)}$	$c_3,c_2,c_5,c_4,c_7,c_8,c_{h(n)}$	0	0	N
8	$3,2,1,4,7,8,h_n$	$p_3,p_2,p_1,p_4,p_7,p_8,p_{h(n)}$	$c_3,c_2,c_1,c_4,c_7,c_8,c_{h(n)}$	0	0	N
9	$3,2,1,4,5,8,h_n$	$p_3,p_2,p_1,p_4,p_5,p_8,p_{h(n)}$	$c_3,c_2,c_1,c_4,c_5,c_8,c_{h(n)}$	0	0	N
10	$3,2,1,4,5,6,h_n$	$p_3,p_2,p_1,p_4,p_5,p_6,p_{h(n)}$	$c_3,c_2,c_1,c_4,c_5,c_6,c_{h(n)}$	0	0	N

Table 3. Routes of $\tau_{3,h(n),2,8}$ and its processing/communication unit set

No.	route	Processing unit set	Communication unit set	$f(\varepsilon_i)$	$f(t_i)$	Available route?
1	$7,8,h_n$	$p_7,p_8,p_{h(n)}$	$c_7,c_8,c_{h(n)}$	0	1	N
2	$7,8,5,6,h_n$	$p_7,p_8,p_5,p_6,p_{h(n)}$	$c_7,c_8,c_5,c_6,c_{h(n)}$	0	0	N
3	$7,8,5,2,3,6,h_n$	$p_7,p_8,p_5,p_2,p_3,p_6,p_{h(n)}$	$c_7,c_8,c_5,c_2,c_3,c_6,c_{h(n)}$	0	0	N
4	$7,8,5,4,1,2,3,6,h_n$	$p_7,p_8,p_5,p_4,p_1,p_2,p_3,$ $p_6,p_{h(n)}$	$c_7,c_8,c_5,c_4,c_1,c_2,c_3,$ $c_6,c_{h(n)}$	0	0	N
5	$7,4,5,8,h_n$	$p_7,p_4,p_5,p_8,p_{h(n)}$	$c_7,c_4,c_5,c_8,c_{h(n)}$	0	0	N
6	$7,4,5,6,h_n$	$p_7,p_4,p_5,p_6,p_{h(n)}$	$c_7,c_4,c_5,c_6,c_{h(n)}$	0	0	N
7	$7,4,5,2,3,6,h_n$	$p_7,p_4,p_5,p_2,p_3,p_6,p_{h(n)}$	$c_7,c_4,c_5,c_2,c_3,c_6,c_{h(n)}$	0	0	N
8	$7,4,1,2,3,6,h_n$	$p_7,p_4,p_1,p_2,p_3,p_6,p_{h(n)}$	$c_7,c_4,c_1,c_2,c_3,c_6,c_{h(n)}$	0	0	N
9	$7,4,1,2,5,6,h_n$	$p_7,p_4,p_1,p_2,p_5,p_6,p_{h(n)}$	$c_7,c_4,c_1,c_2,c_5,c_6,c_{h(n)}$	0	0	N
10	$7,4,1,2,5,8,h_n$	$p_7,p_4,p_1,p_2,p_5,p_8,p_{h(n)}$	$c_7,c_4,c_1,c_2,c_5,c_8,c_{h(n)}$	0	0	N

Table 4. Routes of $\tau_{7,h(n),2,8}$ and its processing/communication unit set

Correspondingly, by (11), (12), (13), and (14), the reliability index of task $\tau_{3,sink,2,16}$ can be deduced. In the same way, the reliability index of each task in Γ' and Γ can be obtained.

5.3. Results

Figure 4 to figure 8 show the reliability index of the tasks vs unit fail rate denoted as *e*. And we have that *e* equals to (1-*r*), where *r* denote the reliability index of the unit.

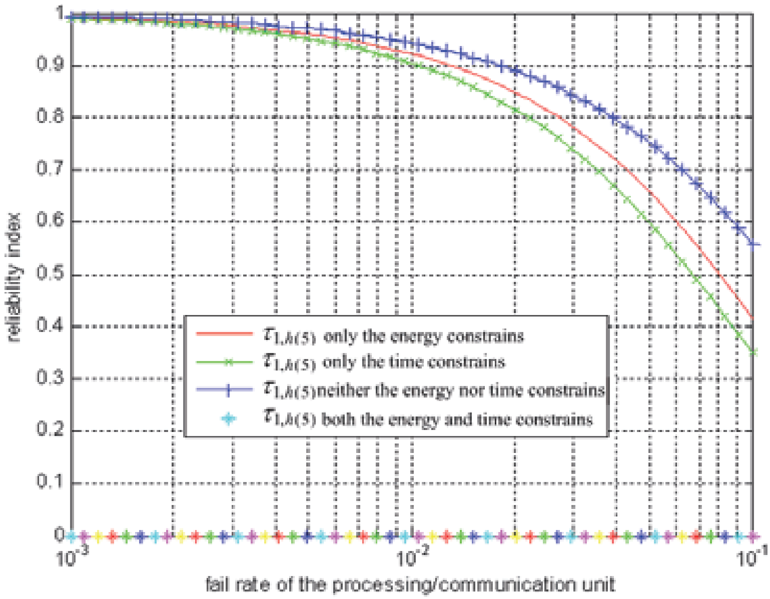


Figure 4. The energy/time constrains' influence on the reliability index of $\tau_{1,h(5),2,8}$

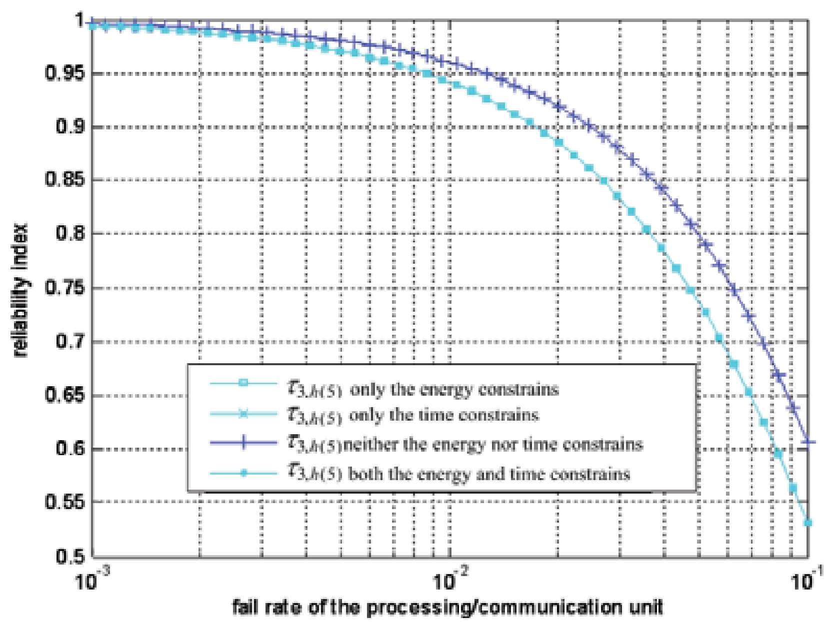


Figure 5. The energy/time constrains' influence on the reliability index of $\tau_{3,h(5),2,8}$

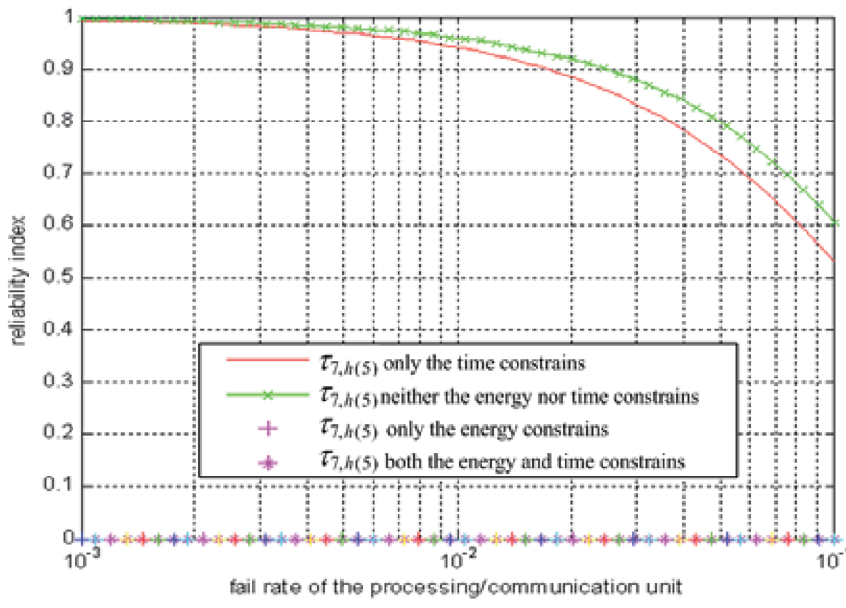


Figure 6. The energy/time constrains' influence on the reliability index of $\tau_{7,h(5),2,8}$

Several results can be obtained directly from these figures, which are as follows:

i. the reliability index of each task decreases with the growth of the fail rate of the processing/communication unit. That means that the reliability index can be increased by improving the reliability index of the unit.

ii. As shown in Figs. 7 and 8, when ignoring the (energy and time) constrains, task $\tau_{3,h(5),2,8}$ or $\tau_{3,sink,2,16}$ has the same reliability index as task $\tau_{7,h(5),2,8}$ or $\tau_{3,sink,2,16}$, respectively. That is concordant with their topology symmetry.

iii. When just considering the energy constrains ($\tau_{S,D,E,\infty}$),

   a. The number of the *energy available* route of Task $\tau_{1,h(n),2,8}$ or $\tau_{3,h(n),2,8}$ is only two. As shown in Figs. 4 and 5, the reliability index of these two tasks has decreased compared with $\tau_{1,h(5),2,\infty}$ or $\tau_{3,h(n),2,\infty}$, respectively. It means that the energy constrains will have a certain extent negative effect on the execution of a task.

   b. There no *energy available* route exists for task $T_{7,h(5),2,8}$, as shown in Fig. 6, therefore, the reliability index of the task is zero. It means that the energy constrains will have a vital effect on the execution of a task.

iv. When just considering the time constrains ($\tau_{S,D,0,T}$),

   a. The number of the *time available* route of Task $\tau_{1,h(n),2,8}$, $\tau_{3,h(n),2,8}$, or $\tau_{3,h(n),2,8}$ is only one. As shown in Figs. 4 ,5, and 6, the reliability index of these three tasks has decreased compared with $\tau_{1,h(5),0,8}$, $\tau_{3,h(5),0,8}$ or $\tau_{7,h(5),0,8}$, respectively. It means that the time constrains will have a certain extent negative effect on the execution of a task.

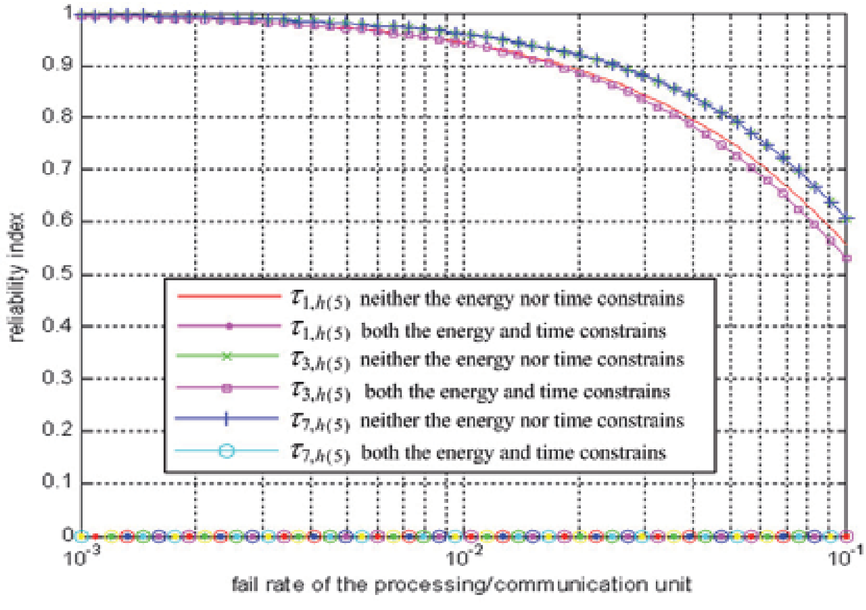


Figure 7. The energy/time constrains' influence on the reliability index of $\tau_{1,h(5),2,8}$, $\tau_{3,h(5),2,8}$, and $\tau_{7,h(5),2,8}$

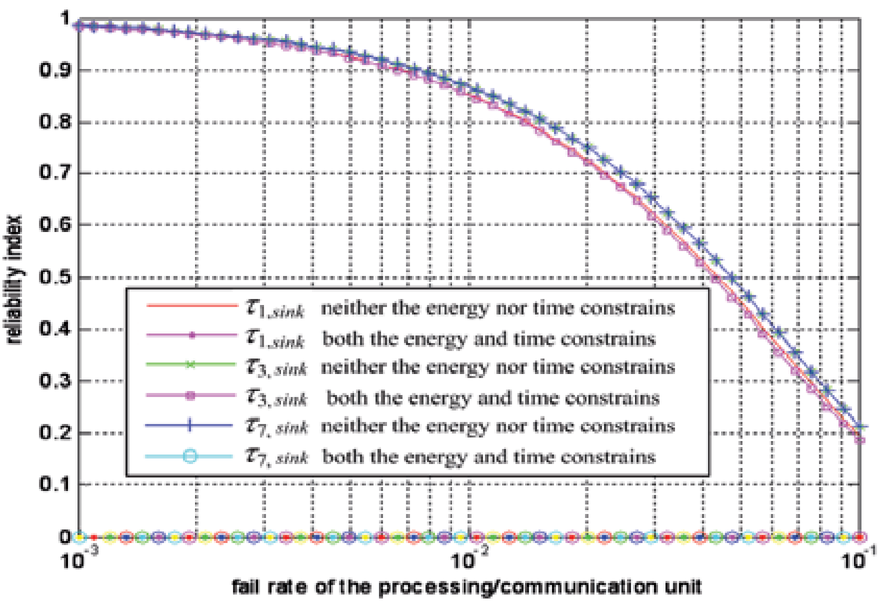


Figure 8. The energy/time constrains' influence on the reliability index of $\tau_{1,sink,2,16}$, $\tau_{3,sink,2,16}$, and $\tau_{7,sink,2,16}$

v. When considering both the energy constrains and the time constrains ($\tau_{S,D,E,T}$),
   a. The number of the *available* route of Task $\tau_{3,h(n),2,8}$ is only one. As shown in Fig. 5, the reliability index of the task has decreased compared with $\tau_{3,h(n),2,\infty}$. It means that both the energy constrains and the time constrains will have a certain extent negative effect on the execution of a task.
   b. There no *available* route exists for task $T_{1,h(5),2,8}$ or $T_{7,h(5),2,8}$, as shown in Figs. 4 and 6, therefore, the reliability index of these two tasks is zero. It means that both the energy constrains and the time constrains will have a vital effect on the execution of a task.

6. Conclusion

The work in this chapter carries on systematically research on unifying energy consume and message delay into reliability modelling for WSN, proposes an integrated model of the task which consider energy consume and message delay for the safety-critical application, introduces both the energy factor function and time factor function, and also establishes an integrated reliability model of WSN based on a task.

The illustration of modelling suggests that the method studied has a directive influence to both task division and topology selection in the phase of system design of WSN system.

Based on this work, future directions can cover several research issues in WSNs:

- To implement reliability validation and optimization for the complicated topologies in WSNs;
- To analyze the reliability index for some kind of topology, search out the key unit (processing unit, or communication unit, or node), and research the redundant scheme for the unit.
- To expand the reliability model to consider more factors (for example, safety or security or buffer limit) than the energy and the time in order to meet the multiple QoS requirements for the safety critical application in WSNs.

Author details

Qiang Zhou
School of Electronic Information Engineering, Beihang University, China

7. References

AboElFotoh, H.M.F.; Iyengar, S.S.; Chakrabarty, K.(2005). Computing Reliability and Message Delay for Cooperative Wireless Distributed Sensor Networks Subject to Random Failures, *IEEE TRANSACTIONS ON RELIABILITY*, Vol.54, No.1, (2005), pp. 145-155

Akyildiz, I.F. ; Su, W. etc. (2002). Wireless Sensor Networks: A Survey, *Computer Networks*, Vol.38, No,4, (2002), pp. 393-422

Banerjee, S.; Khuller, S. (2001). A Clustering Scheme for Hierarchical Control in Multi-hop Wireless Networks. *IEEE INFOCOM 2001* , Anchorage, Alaska, April 2001.

Callaway, E. H. (2004). Wireless Sensor networks: Architectures and Protocols. *Auerbach Publications,* Vol.12, No.6, (2004), pp.88-90

Edgar, H.; Callaway, J.; Callaway, E. H. (2003). *Wireless Sensor Networks: Architectures and Protocols*, CRC Press

Feng, Xue; Kumar, P.R. (2004). The Number of Neighbors Needed for Connectivity of Wireless Networks, *Wireless Networks,* Vol.10, No.2, (2004), pp. 169-181.

Heinzelman, W.; Chandrakasan, A.; Balakrishnan, H. (2000). Energy-Efficient Communicat ion Protocol for Wireless Microsensor Networks. *In Proceeding of the 33rd Hawaii International Conference System Sciences* , vol. 8, IEEE, Maui, Hawaii, Jan 2000.

Kim, T.; Lee, Y. ; Sung J. ; et al. (2010). Hierarchical Network Protocol for Large Scale Wireless Sensor Networks, *Proceedings of The 7th IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 1-2, 2010

Johannes U.H.; Soh S.; Rai S.; et al. (2012). Computing Performability for Wireless Sensor Networks. *International Journal of Performability Engineering,* Vol.8, No.2, (2012), pp.131-140.

Lin, P.M. (1976). A New Algorithm for Symbolic System Analysis. *IEEE Trans on Reliability,* Vol.25, No.1, (1976), pp.1-15.

Silva, I.; Guedes, L.A.; Portugal, P.; et al. (2012). On Reliability and Availability Evaluation of Wireless Sensor Networks for Industrial Applications. *Sensors,* No.12, (2012), pp. 806-838.

Sun, L.; Li, J.; Chen, Y. (2005). *Wireless sensor networks*, Tsinghua University Press, Beijing, China

Tilak, S.; Abu-Ghazaleh, N.B.; Heinzelman, W.(2002). A taxonomy of wireless micro-sensor network models. *ACM Mobile Computing and Communications Review,* Vol.6, No.2, (2002), pp.28–36

Tripath, C.R. (1997). Reliability Analysis of Hypercube Multi-computers. *Microelectron Reliability*, Vol.37, No.6, (1997), pp. 885-891.

Tubaishat, M.; Madria, S. (2003). Sensor Networks: An Overview. *IEEE Potentials,* Vol.22, No.2, (2003), pp.20-23

Tubaishat, M.; Yin, J.; Panja, B.; Madria, S. (2004). A Secure Hierarchical Model for Sensor Networks. *ACM SIGMOD,* Vol.33, No.1, (2004), pp.7-13

Xing, Liudong; Shrestha, Akhilesh (2006). Qos Reliability of Hierarchical Clustered Wireless Sensor Networks, *Proceedings of The 25th IEEE Intl. Performance Computing and Communications Conference*, pp. 641-646, 2006

Xing, L.(2006). Integrated Modeling for Fault-Tolerant Sensor Networks Reliability and Security, *The 52nd Annual Reliability & Maintainability Symposium (RAMS06),* Newport Beach, CA, January, 2006

Zhang, G. D.;Lu, X. T.; et al. (1997). *Design and analysis of system reliability and maintenance,* National Defence Industry Press, Beijing, China

Zhou, Q.; Li, Q.; Luo, Z.Q.; et al. (2001). On Reliability Model of SCI Interconnections in Avionics System. *Journal of Beijing University of Aeronautics and Astronautics,* Vol.27, No.5, (2001), pp. 536-539.

A Framework for Integrating Wireless Sensor Networks and the Internet

Jeisa Domingues, Antonio Damaso, Rilter Nascimento and Nelson Rosa

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50118

1. Introduction

The integration of Wireless Sensor Networks and the Internet is growing in importance as WSNs have been employed in a great variety of applications which need a common means to share data over the Internet. In the recent years, many solutions have been proposed to provide that integration. The simplest and most popular one is the gateway-based approach [1], which basically converts between protocol stacks and logical address formats used in both networks. In the overlay-based approach [2], some sensor nodes may implement the TCP/IP protocols or some hosts may implement WSN protocols, being the reachable nodes and acting like gateways in their respective network. Another solutions use the TCP/IP suite as the communication protocols for sensor nodes [3]. These three approaches focus on accessing the network nodes through their logical addresses, which has several problems, such as the different addressing and routing scheme of both networks. Besides, employing TCP/IP in the sensor nodes raises specific issues, like the header overhead of those protocols which is very large for small packets, and the end-to-end retransmissions used by TCP which consume energy at every hop of the path.

Mobile agents have also been used as an approach to dynamically access the WSN from the Internet [4], answering queries while migrating through the sensor nodes. Another solution adopts service-oriented middleware to integrate WSNs and the Internet, converting sensor nodes into service providers for the Internet hosts [5]. Additionally, Web service approaches have been presented [6], some of them converting all the WSN into a single web service, and others allowing sensor nodes to offer their data through Web services that can be accessed from the Internet. The main limitation of these approaches is that sensor nodes are considered only as service providers, not consumers.

In order to overcome the aforementioned problems, this chapter introduces a framework for integrating WSNs and the Internet by allowing the interoperability of their services. Our approach aims at integrating applications (considered as services) instead of networks (i.e., protocol stack and/or logical address formats mapping). This service abstraction is an important advantage, since it offers to the application developers an easy and transparent way to integrate the networks with seamless handling of addresses and protocols. Furthermore, the provision of the capability of transparently requesting services in both directions is another benefit offered by our approach. Although sensors are typically providers of sensed data, there are some cases where it is better for the sensor to request a service outside the WSN. That happens, for instance, when the application is too computationally demanding, when it needs to store a huge amount of data, and/or when it requires global knowledge of the WSN.

The rest of this chapter is organized as follows. Section 2 describes a scenario of usage and an overview of the framework. The framework elements, namely Smart, WISeMid, SAGe and Clever, are presented in Sections 3, 4, 5 and 6, respectively. Section 7 discusses an energy consumption evaluation of those elements, and a real application that emphasizes the usefulness of the sensor nodes' capability of being service requestors is also analyzed. Finally, Section 8 presents some concluding remarks.

2. Illustrative scenario

This section presents a scenario of usage of the framework which is depicted in Figure 1 and comprises two distributed applications.

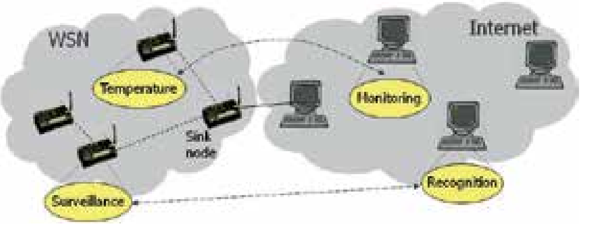


Figure 1. Scenario of usage of the framework

The first application monitors an area (such as a house, a factory, etc.) to detect and report the presence of intruders. It is composed by two services: Surveillance and Recognition. The Surveillance service runs in a sensor node and captures images of any moving target. Each captured image has to be analyzed to verify if the moving target is really a potential threat. This is necessary to avoid reporting the presence of a cat, for example. Therefore, the application has to perform an image recognition procedure to classify the moving target as human or non-human. This kind of task is too computationally demanding and resource consuming, which makes it non-suitable for the resource constraints of the sensor nodes. Considering this, it may be more worthwhile for the Surveillance service to invoke an image recognition service outside the WSN instead of performing itself that task.

Now consider that an Internet host provides the (image) Recognition service, which offers an operation that receives an image, analyzes it and classifies it as human or non-human. The Surveillance service sends the captured image to the Recognition service and, if it signals that the target is a human, the Surveillance service reports the presence of the intruder.

The second distributed application monitors the temperature of an area (such as a forest) to detect high temperatures and prevent potential fires. This application is composed by two services: Monitoring and Temperature (see Figure 1). The Monitoring service runs in an Internet host and keeps track of the environment temperature of the area of interest. For that purpose, it uses a Temperature service that runs in a WSN which has been deployed in that area. This service measures and returns the environment temperature value. Hence, to observe the changes in temperature, the Monitoring service calls the Temperature service periodically.

Both described applications spread out through the Internet and a WSN. In order to allow the construction of this kind of distributed application composed by services from both networks, some issues should be addressed. First, there must be a uniform way of describing those services regardless of their provider being sensor nodes or Internet hosts. Then, based on that description, there must be a mechanism to discover those services and use them. Those services should be able to communicate with each other irrespective of their location. Also, the service location should not influence the way it is accessed. In other words, a service provided by a WSN node should be accessed the same way that a service provided by an Internet host.

Additionally, the distributed applications described above focus on specific areas of interest. The surveillance, for instance, would define the border region of a WSN as a critical one to be monitored, whereas forest fire detecting applications may define clearings that resulted from previous fires as the critical regions to be monitored (since they tend to be more propitious for new fires due to their dry vegetation and soil). Therefore, there should be a mechanism to select a group of nodes in the WSN so such areas can be delimited.

Those issues are handled by the presented framework as described in the following sections.

2.1. Framework overview

For the purpose of enabling the construction of applications that are distributed between WSN's and Internet's nodes, the framework comprises four components: a service model to describe services from both networks; a communication infrastructure (middleware) that enables the interoperability of WSN's and Internet's services; an advanced gateway which is responsible for providing the location and access transparency for the services' communication; and a service composition tool which enables the creation of logical regions in the WSN by just composing Web services that are available on the Internet.

The service model is named Smart (**S**ervice **m**odel for integr**a**ting Wi**r**eless Sensor Networks and the In**t**ernet) [7] and is able to describe both WSN's and Internet's services. Describing a service in a uniform way despite its location is essential to provide the integration of those networks at service level. The Smart model allows a service provider to characterize its service by making available some information about it. That information includes functional details, such as the service interface, nonfunctional details, such as service properties, and information on how to interact with it so the service can be accessed.

Once the services have been described, a middleware called WISeMid (**W**ireless sensor network's and **I**nternet's **S**ervices int**e**gration **Mid**dleware) [8] provides a communication infrastructure for the services interaction, supporting the integration of WSNs and the Internet at service level. In this context, applications running in the Internet/WSN nodes may play the role of service providers or service users, where a service user is able to communicate with a service provider no matter whether they are running in the same network or not. For that purpose, WISeMid provides an infrastructure that allows integrating these services in such a transparent manner that services provided by both WSN nodes or Internet hosts are accessed the same way. Moreover, WISeMid implements the following mechanisms for saving energy in WSNs [9]: Aggregation service, which aggregates the last n data sensed by a node; Reply Storage Timeout, which avoids sending equivalent messages to the sensor nodes while the last sensed data is still considered up-to-date; Automatic Type Conversion, which removes unnecessary bytes from the messages; and the implementation of asynchronous invocation patterns, which prevents the sensor application from wasting power for being blocked during a service request.

The access and location transparency provided by the middleware is performed by the gateway, which is called SAGe (**S**ensor **A**dvanced **G**at**e**way) [10]. Running in an Internet host which is connected to the WSN sink node, SAGe's main function is to act as a service proxy between both networks by enabling the communication between services running on Internet hosts and WSN nodes in such way that the service user should invoke a service without knowing if it is being provided by a WSN node or an Internet host. Additionally, SAGe is involved in all energy-saving methods performed by WISeMid.

The service-level integration of the WSNs and the Internet is enhanced by a Web service composition tool named Clever (**C**omposing **l**ogical r**e**gions **v**ia s**er**vices). This tool enables the definition of logical regions in WSN by composing Web services available in the Internet. Having a Web service acting as a proxy to each WSN node, the definition of a logical region in the WSN consists of defining a Web service composition. Once that composition is deployed, a new Web service is created and an invocation to it is passed to all sensors belonging to the WSN logical region and represented by the composition.

The next sections will describe in details each of those components.

3. Smart

The first element of the framework is Smart (**S**ervice **m**odel for integr**a**ting Wi**r**eless Sensor Networks and the In**t**ernet), a service model that is suitable for both Internet and WSN services, promoting the integration of those networks. In other words, both services designed for a WSN and services designed for the Internet may be described using the Smart model. Although there are many service models for the Internet [11, 12] and a few service-oriented models for WSNs [13, 14], for the best of our knowledge, Smart is the first service model able to describe both WSN and Internet services.

Either on the Internet or in a WSN, a service provider needs to characterize its service by making available some information about it. The essential and simplest information that must be available is the service identification and what it can do. The service interface offers that information by specifying the service name and the operations it provides. A description of how the service provides its operations may also be exposed. That information concerns the service functionality and thus compose the service functional description.

A service user may be interested in knowing specific details of the service provider, the service implementation or its running environment. That information is called nonfunctional description and may include, for instance, the bit rate of a service provider or its location.

Furthermore, for a service to be accessed, information on how to interact with it needs to be defined. That is so called interoperability information and includes the format of the messages the service understands and its access point.

With that in mind, the Smart model defines a service as being composed by functional and nonfunctional descriptions, from which depends its interoperability information.

3.1. Functional description

To be able to request a service and the execution of one of its operations, a service user (client) needs first to be aware of the service name and what operations it can perform. More than that, the client needs to know the input and output data of a successfully execution of the service, as well as the possible errors that an unsuccessful invocation may generate in order to handle them. That information is commonly referred to as the service interface.

Some services share the name and the operations they provide, but may produce different results by performing different actions or being under different conditions to execute the same operation. Those services are said to have the same type, and the differences between them are expressed using properties, which assume different values from service to service.

Another service characteristic that concerns its functionality is the period of time the service instance needs to exist, which is known as its lifecycle and is closely related to maintaining the service state between subsequent invocations. According to the lifecycle patterns defined in [15], service instances may be static, per-request or per-client. A static instance may be used when the service is independent of any service user (client), and its state must be available to all clients between individual invocations. When the service does not require maintaining state and is accessed by many users at the same time, an instance per request is more suitable. A per client instance is appropriate when it is important to maintain the service state for subsequent invocations by the same user, which happens when the logic of the service extends the logic of the client.

Considering all that information about the service functionality, the Smart model defines the service `FunctionalDescription` as being composed by its `Interface` and its `Behavior`, as shown by the class diagram in Figure 2. Those elements are explained in detail as follows.

3.1.1. Interface

As previously stated, the interface contains the service name and the operation(s) it provides (see Figure 2). The service's `name` is used by a service client to locate a particular service.

Concerning the service operations, each `Operation` is defined by a `name`, an operation `type`, a list of typed parameters (`Argument`), a result (`Outcome`) and a list of possible errors (`Fault`). There are two types of operation: `Interrogation`, which is a request-response operation; and `Announcement`, which is an one-way operation. The list of errors represents exceptions raised by the operation, each of which is defined by a `name` and a `message`.

In the WSN context, an operation is either a `Command` or an `Event`, which is actually a feature of nesC, one of the most used programming language for developing applications for WSNs [16]. Commands are implemented by the interface's provider, whereas events are implemented by the interface's user.

The `ServiceType` is used to categorize services according to their capabilities. That means that all services of the same type, share the same `interface` and the same `properties`,

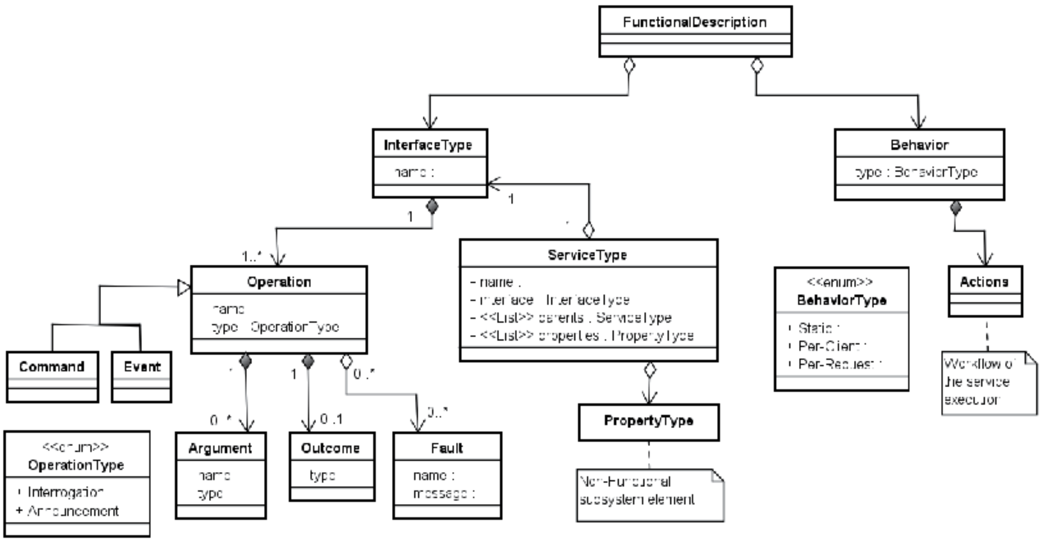


Figure 2. UML Class diagram of the Functional description

being distinguished by the different values of their properties. Also, a service type may extend another one (`parent`), inheriting its interface signature, while being able to add more operations, and its properties, also being able to define additional ones.

3.1.2. Behavior

The service behavior includes a `BehaviorType` and a set of `Actions`, as depicted in Figure 2. The behavior type defines the service's lifecycle: static, which specifies that the service has a unique instance; per-client, which states the service has an instance for each client; and per-request, meaning that each request is handled by a different instance of the service.

The `Actions` element defines the service behavior in terms of activity structure. It specifies the workflow (execution) of the service. This workflow may be simple, specifying the expected execution order of the service operations, which is necessary when the execution of an operation affects the execution of other one. It may also be more complex, identifying the actions of the operation and specifying the ordered sequence in which they are performed.

3.2. Nonfunctional description

The functional description specifies *what* the service can do, but describing *how* the service can do it may be very useful or even necessary for a service user to choose the service that best fits their needs. This extra description includes characteristics (properties) that are not directly related to the functional description of the service, such as the server geographic location or the service security, and therefore is commonly referred to as nonfunctional description.

A particular subset of nonfunctional characteristics is very important for choosing the proper service: quality of service (QoS) parameters. Although the term QoS is typically related to network performance parameters (such as bandwidth, latency and error rate), in a service related context it covers a wider range of service properties which can be used as quality indicators, including, for instance, accuracy, dependability, robustness, security, customer service, etc.

With that in mind, Smart defines that the `NonfunctionalDescription` of a service specifies its nonfunctional properties (`Property`) and QoS parameters (`QoSParameter`), as depicted in Figure 3. Those elements are explained in detail next.

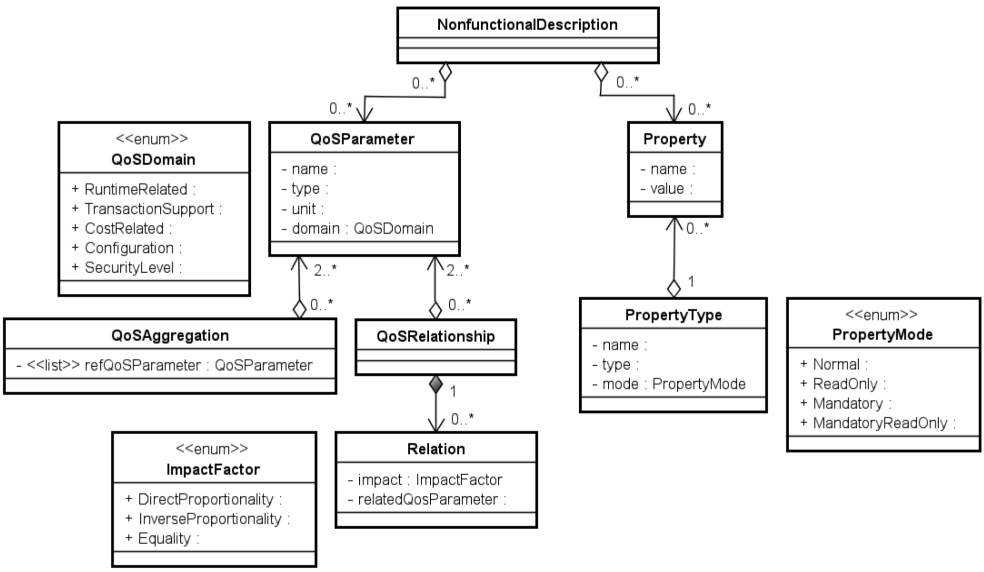


Figure 3. UML Class diagram of Nonfunctional description

3.2.1. Property

The Smart model elements which describe properties (`Property`, `PropertyType` and `PropertyMode`), are based on the CORBA Trading Object Service Specification [17].

In Smart, each `Property` can be described by a name-value tuple, where `name` is a string that names the property and `value` is the value assigned to the property (see Figure 3).

The service properties are actually classified in property types, which are related to the `ServiceType` element of the functional description (see Figure 2). A `PropertyType` is defined by a `name`, a property value `type` (e.g., integer, string, etc.) and a `mode`. The mode (`PropertyMode`) defines whether a property is mandatory and/or readonly. When a property is `Mandatory`, it means that an instance of the service type that defines this property must provide an appropriate value for it when registering the service. When it is `ReadOnly`, the value for this property is optional, but if provided, it may not be changed after the service is registered. When it is both mandatory and readonly (`MandatoryReadOnly`), the property value should be provided when registering the service and may not be changed after that. Finally, when a property is `Normal`, it is optional and its value may be subsequently modified.

3.2.2. QoS Parameters

QoS parameters refer to service properties that can be used as indicators of quality, as said previously. Based on the SMEPP service model description [18], the Smart model defines that a `QoSParameter` is described by a `name`, a `type`, a `unit` and a `QoSDomain` (see Figure 3). The `name` is a string that specifies the QoS parameter name, such as "throughput". The `type` describes the data type of the QoS parameter value (e.g.: integer, float, etc.), whereas the `unit` represents the unit of measurement (such as second, bits, percent, etc.).

A `QoSDomain` specifies the domain of the information enclosed by the QoS parameter. Possible QoS domains include runtime (e.g.: performance), transaction support (e.g.: integrity), configuration and cost (e.g.: stability), and security (e.g.: authentication) [19].

Some QoS parameters are composed by two or more parameters. For example, the "performance" parameter is composed by "response time", "throughput" and "latency". Such compound QoS parameters may be described by a `QoSAggregation`, with two or more QoS parameters, each of which represents a parameter member of the composition.

Also, QoS parameters may affect other ones and this influence is expressed by the `QoSRelationship`, which defines the impact factor as one of the following values: `DirectProportionality` (i.e., the related parameter values have the same behavior: if one increases or decreases, so does the other), `InverseProportionality` (i.e., the parameter values have the opposite behavior: if one increases, the other decreases; and vice-versa) or `Equality` (i.e., the parameters have the same value).

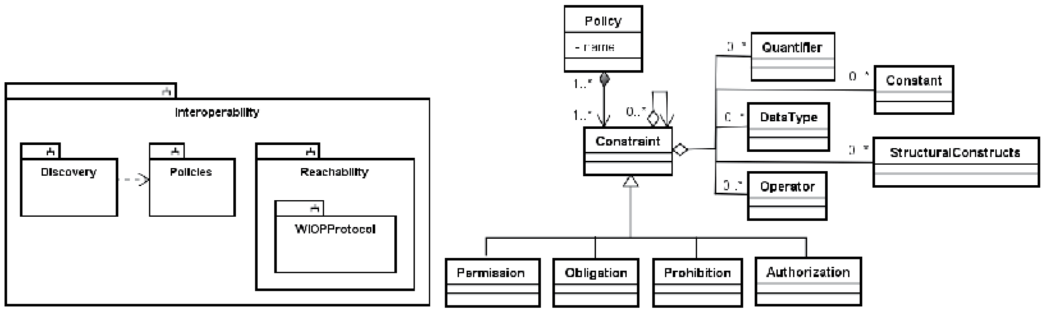
3.3. Interoperability information

According to [20], interoperability *"is about the meaningful sharing of functionality and information that leads to the achievement of a common goal"*. For that to be possible, a message exchange pattern should be defined by the model. It specifies the message sequence and the exchange rules, which should be followed by both sides of the interaction and is usually specified by a protocol definition. Also, for the interoperability to happen, the services must provide a kind of identification, a way to be accessed.

Before being able to exchange messages, the services need to find each other. By finding, we mean searching and retrieving information about the services which satisfy some user defined conditions. The service discovery procedure is responsible for that task and is implemented by some special services that offer operations for registering and searching services. The most simple discovery service is called Naming service and locates a service based only on the service name. Trading service is another special service that implements a discovery procedure, which discovers a service based on its type and a set of property values that describe it.

Still concerning service interoperability, a set of policies may be defined to guide the services interaction. Different policies may be involved in the process of service discovery to check the constraints that either the service provider or the service user specify to protect the process of communication or sharing the information.

In Smart, all that information that concerns the way a service may interact with another is represented by `Interoperability` element, which consists of `Policies`, `Discovery` and `Reachability`, as illustrated in Figure 4(a).



(a) UML Subsystem diagram of Interoperability information

(b) UML Class diagram of Policy

Figure 4. UML diagrams of Interoperability information

3.3.1. Policy

As stated previously, policies may be used by either the service provider or the service user to protect the process of communication or sharing the information. For that end, each `Policy` defined by a service prescribes a set of `Constraints` for interacting with it, as depicted in Figure 4(b). Those constraints define the ideal, acceptable or desirable service behavior during an interaction with another service.

In the Smart model, a `Constraint` may describe an `Obligation` – a behavior that is required; a `Permission` – a behavior that is allowed to occur; a `Prohibition` – a behavior that must not occur; and an `Authorization` – a behavior that must not be prevented for the services involved. Note that a permission is equivalent to there being no obligation for the behavior not to occur, whilst an authorization is actually an empowerment.

Constraints may contain quantifiers (e.g.: for all, there exists), constants, data types (e.g.: integer, String), operators (arithmetic operators, comparison operators, boolean operators, implication operators) and structural constructs (e.g.: let in, if then else) [21].

3.3.2. ReachabilityInfo

The `ReachabilityInfo` element is in charge of providing the service information that enables other services to locate and interact with it. Therefore, it should define the `EndPoint` to which a client can direct messages to invoke actions and the `MessageExchangePattern`, which specifies the protocol to adopt for message exchange using the endpoint (see Figure 5).

Since Smart aims at integrating the Internet and WSNs, it specifies two kinds of end point to a service, one for each network. For an Internet service, the `InternetEndPoint` is described by the IP address of the host (`ipAddress`) and the port number through which the service is provided (`portNumber`). For a WSN service, the `WsnEndPoint` may identify a `nodeId` or a `groupId`, as a service may be provided by a single node or by a group of nodes.

For the `MessageExchangePattern`, the Smart model defines that the `Publish/Subscribe` communication model or the `Request/Reply` may be used. The former has been chosen due to its suitability in WSNs, since it saves energy by sending a message only when an event of interest is detected. The request/reply approach has been chosen for being the most suitable for server/client communication. For now, only the

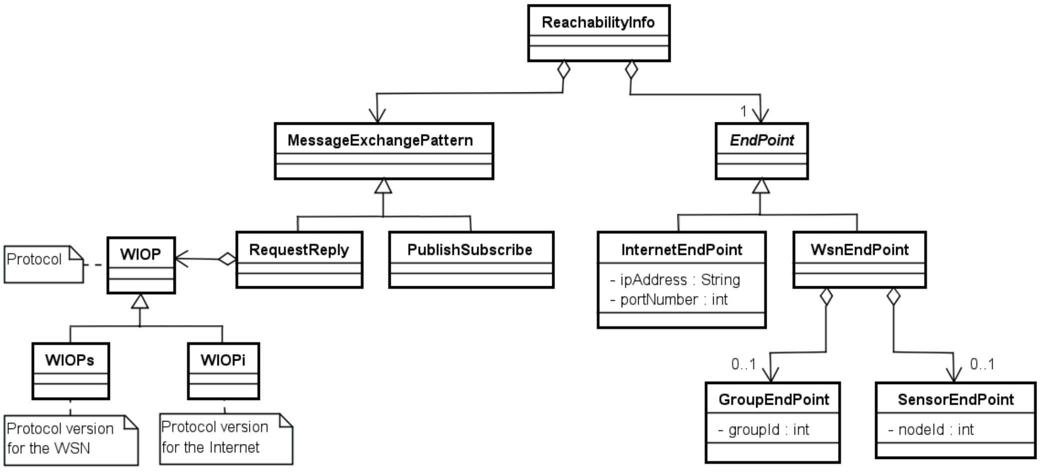


Figure 5. UML Class diagram of Reachability information

request/reply has been detailed in the model, with a protocol being defined for that purpose. The request/reply protocol, namely `WIOP` (see Section 4.3.1 for details), defines different message formats for Internet service (`WIOPi`) and WSN services (`WIOPs`), to handle their specificities, and provides the mapping among those formats (especially concerning the data types) in order to ensure that the exchanged information has the same meaning for both message sender and receiver, no matter if one is located at the Internet and another at a WSN.

3.3.3. DiscoveryProcedure

As its name suggests, the `DiscoveryProcedure` element models the service discovery procedure, which may be defined as the act of finding a service that may have been previously unknown and that meets certain functional criteria. Since additional nonfunctional criteria may be used to locate the desired service, the `DiscoveryProcedure` element interacts with the FunctionalDescription and the NonfunctionalDescription elements, as depicted in Figure 6.

Smart defines two discovery procedures: the `Naming`, that uses only functional criteria, and `Trader`, that uses functional and nonfunctional criteria. Both procedures are detailed next.

3.3.3.1. Naming

The `Naming` component represents the Naming discovery procedure, which is used to register/locate a service based on its name. To provide scalability to the Naming Service, the Smart model defines two architectures for it: `Flat` and `Hierarchical` (see Figure 6). The flat Naming uses unique, globally distinguished names, whilst the hierarchical one deals with different domains. Therefore, the chosen architecture defines the name space, that comprises a `Context` in which names are unique and valid.

3.3.3.2. Trader

The `Trader` component represents the Trading discovery procedure. The Smart model definition of the trading function is based on the CORBA Trading Object Service Specification [17] and the ANSA Model for Trading and Federation [22].

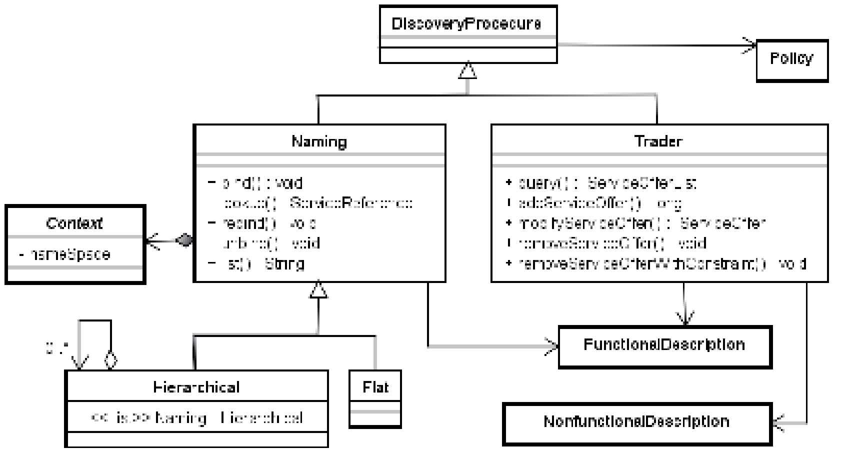
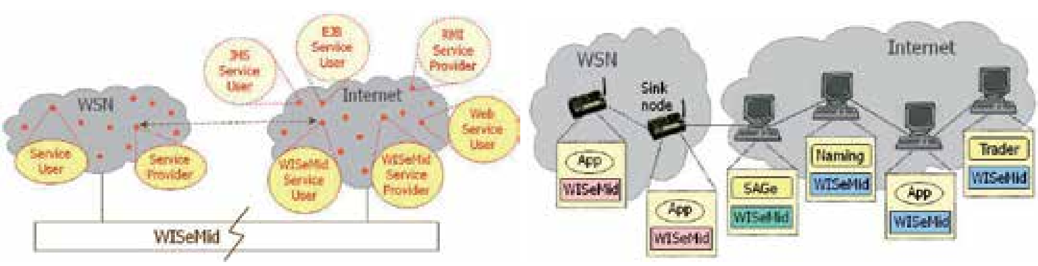


Figure 6. UML Class diagram of Discovery

As stated before, the trading service is concerned with discovering a service based on its type and a set of property values that describe it. The act of registering information about a service in the Trader is usually called exporting a service offer. A service offer consists of the service type name, a reference to the interface that is provided by the service, and zero or more property values for the service. Those properties correspond to the ones that are listed in the associated service type (and parents), and specifying a value for them depends on the mode assigned to each one of them. A value must be provided for every property that has been defined as mandatory (including both Mandatory and MandatoryReadOnly modes) in the service type. Providing a value for the other properties is optional.

4. WISeMid

As mentioned earlier, WISeMid provides a communication infrastructure for services interaction, where applications running in the Internet/WSN nodes act as service providers or service users that are able to communicate irrespective of being at the same network or not. As depicted in Figure 7(a), those applications (services) should be developed in different technologies, such as Web Service, Java RMI, EJB and JMS, although WISeMid current implementation supports only WISeMid services and Web services.



(a) WISeMid logical view (b) WISeMid physical view

Figure 7. WISeMid views

Figure 7(b) presents a physical view of how WISeMid spreads out through the Internet and WSN. The WISeMid implementation for WSN and Internet is not the same, as they have different requirements and components (that will be explained in the subsequent sections). Physical communication is performed through an Internet host that is connected to the WSN sink node via a serial port (USB). This host executes a special WISeMid service called SAGe, which acts as a proxy between both networks (see Section 5).

4.1. WIDL

The first step to enable the services interaction is to specify a notation to describe a service. For this particular purpose, we have defined the WISeMid IDL – WISeMid Interface Definition Language. It enables us to define service interfaces in a uniform way, wherever the service runs (Internet or WSN) and whatever the implementation language (Java or nesC). Based on Smart model definition of the service interface and its properties (see Sections 3.1.1 and 3.2.1), the general structure of a service definition in WISeMid IDL is presented below:

```
1: module PACKAGE_NAME{
2:   interface INTERFACE_NAME{
3:     [OPER_TYPE] OUTCOME_TYPE  OPER_NAME([TYPE ARG1,...]) [raises(EXCEPTION_NAME1,...)]
4:   }
5:   [type TYPE_NAME [extends TYPE_NAME] {
6:     propertyType [PROP_MODE ,] PROP_NAME , PROP_TYPE;
7:   }]
8:   [properties TYPE_NAME {
9:     [property PROP_NAME = VALUE;]
10:  }]
11:}
```

First, the module (package) that contains the service should be specified (1), followed by the service interface, which includes its name (2) and provided operations (3). Each operation has a name, input/output typed parameters and may raise exceptions. An operation is by default a request-response operation, but it may be defined as a one-way operation by using the reserved word `oneway` as the operation type. (Note that in Smart, request-response corresponds to the `interrogation` type and one-way is the `announcement` type.) The service type definition (5-7) includes its name, its parent's name (5), and a list of property types, each of which comprising a property mode (optional), a name and a type (6). The property mode options are `NORMAL` (default), `MANDATORY`, `READ_ONLY` and `MANDATORY_READ_ONLY`.

Property values are defined in `properties` clauses (8-10), which are related to the service type that specified the property type (8). Every property type defined as `MANDATORY` or `MANDATORY_READ_ONLY` in the service type must have a property value associated to its name in the `properties` clause (9). Specifying a value for property types of other modes is optional. Additionally, property values that have not been defined in the service type may be specified. The definition of service type and properties is optional.

The service definition in a WIDL file may be used for automatic code generation. To perform that task, we have developed a tool, namely ProxiesGen, that can interpret a WIDL file and generate the proxies (i.e., stub and skeleton) related to the described service according to a command line parameter, which specifies if the service is supposed to be provided by a WSN node or by an Internet host. When service type and properties are defined, ProxiesGen adds to the server code the creation and registration of the service type in the Trader. Also, a service offer with the defined property values is created and exported.

4.2. Architecture

Based on the DOC middleware layers [23], the WISeMid architecture consists of three layers: Infrastructure, Distribution and Common Services (see Figure 8).

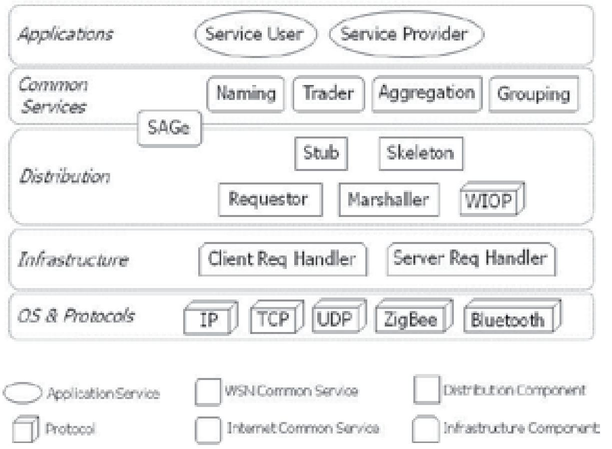


Figure 8. WISeMid Architecture

The Common Services layer includes services that are not particular to a specific application domain: Aggregation, which performs sensor data aggregation; Grouping, which defines clusters inside the WSN; Naming and Trader, that store information needed to find and access a service (they run in the Internet); and SAGe, that is in charge of converting and forwarding messages from/to WSN (it runs in the Internet). Additionally, SAGe also provides location transparency acting as a service proxy between both networks, as described in Section 5.

The Distribution layer includes the following elements: Stub, Requestor, Skeleton and Marshaller, which are basic middleware elements [15]. Those elements handle WIOP messages. WIOP is the request/reply protocol specified by Smart (see Section 3.3.2).

The Infrastructure layer consists of the Client Request Handler and the Server Request Handler, which handle network communication using the communication facilities provided by the operating systems, e.g., sockets (Windows) and ActiveMessageC (TinyOS).

4.3. Implementation

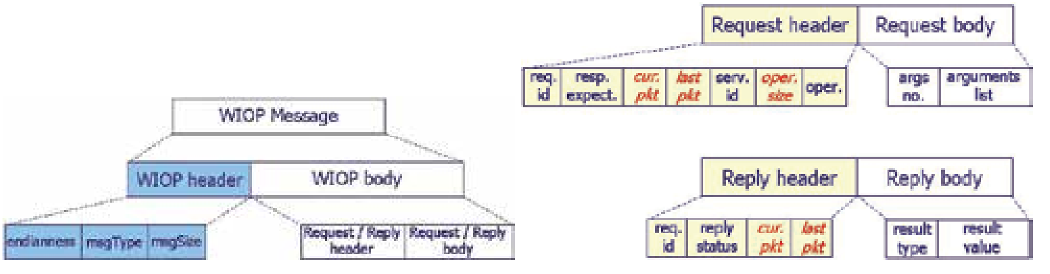
The WISeMid implementation is divided into two parts, one for the WSN nodes, developed in nesC, and another for Internet hosts, developed in Java. The elements of the infrastructure layer (Client Request Handler and Server Request Handler) and of the distribution layer (Stub, Skeleton, Requestor, Marshaller and WIOP) have been implemented in both languages, as nodes from both networks may play the role of user or provider of the service. Actually, due to sensor nodes' limited resources, some elements are not present in the WSN: the Requestor is not implemented by the WSN service users (its functions are deployed by the Stub), and WIOP messages are treated as byte sequences, making the Marshaller unnecessary.

The above elements, except for WIOP, are basic middleware elements, thus their implementation will not be described here. The Aggregation and the Grouping elements are common services that run in the WSN. The Aggregation service may be considered as a method to save energy in the sensor nodes and therefore it is described in Section 4.3.4 with other energy-saving approaches that have been implemented in WISeMid. The Grouping service was implemented as a web service composition approach that enables grouping the WSN nodes into logical regions. This approach is used by Clever and described in Section 6.1.

As SAGe has an essential role in WISeMid, it is described in Section 5. The other architecture elements (WIOP, Naming and Trader) are described as follows.

4.3.1. WIOP

Specified by the Smart model as the message exchange pattern to enable service interoperability (see Section 3.3.2), the WISeMid Inter-ORB Protocol (WIOP) is a GIOP-based protocol that defines the Request/Reply messages between clients and servers. A WIOP message is divided into header and body, as depicted in Figure 9(a).



(a) WIOP Message header (b) WIOP Message body

Figure 9. WIOP (WISeMid Inter-ORB Protocol)

The WIOP message header is composed by the following fields: `endianness` (e.g., big endian or little endian); `msgType` (e.g., request or reply message); and `msgSize`, which stores the message size in bytes. It is worth noting that WIOP messages do not contain fields for source/destination ID (such as sensor ID or IP address). That happens because, as other inter-orb protocols (e.g. GIOP), WIOP uses the protocols of the lower layers in both networks (that is, Active Message in WSN and TCP/IP in the Internet), and WISeMid infrastructure, more specifically SAGe, handles the information concerning addressing.

The WIOP message body may contain a Request or a Reply message. Those messages, which have also a header and a body, are illustrated in Figure 9(b).

The Request message header's main fields are: `requestId`, which stores the Request message ID; `responseExpected`, which signals whether the request expects a Reply message or not; `serviceId`, which is the ID of the requested service; and `operation`, which represents the name of the operation being invoked. Those are the fields that compose the Request message header in the Internet version of WIOP (called $WIOP_i$). The version that runs in the WSN (called $WIOP_s$) contains three additional fields (emphasized in Figure 9(b)):

`currentPacket` and `lastPacket`, which are the current and last packet ID, respectively, and `operationSize`, which represents the length of the operation name. This last field is necessary because the operation field has no fixed size. It is adjusted to the length of the operation name, to keep the message as minimal as possible. Furthermore, a single $WIOP_s$ may not be enough to transmit the total amount of data to/from a WSN service, due to the size limitation imposed by TinyOS' Active Message [24]. In such cases, the data is divided in many $WIOP_s$ messages, and the `currentPacket` and `lastPacket` fields are used to control the message fragmentation/reassembly at the sensor node and SAGe.

The Request body consists of the number of arguments (`numArgs`) followed by a sequence of type and value of each argument. Also, the way arguments are stored in the Request body is different for the WSN version. In the $WIOP_i$, the arguments are stored one by one, being each argument composed by a type and a value. For example, three arguments would be stored in the following sequence: type1, value1, type2, value2, type3, value3. In the $WIOP_s$, only the first argument is individually stored (with its type and value in a row). From the second argument on, the arguments are grouped into pairs where the types of both arguments come first followed by their respective values. That happens because types are represented by integer numbers between 0 and 11, and therefore each type can be stored in only 4 bits. Hence two types can be grouped into one byte, being followed by their related argument values. The first type is a special case as it uses the second half of the byte that carries the `numArgs` field. Considering this configuration, three arguments would be stored in the following sequence: type1, value1, type2, type3, value2, value3.

The Reply message header main fields are: `requestId`, which stores the related Request message ID; and `replyStatus`, which signals whether there was any exception while executing the request, and its possible values are NO_EXCEPTION (0), USER_EXCEPTION (1), SYSTEM_EXCEPTION (2) and LOCATION_FORWARD (3). As the Request message header, the Reply message header in the $WIOP_s$ contains the additional fields related to the message fragmentation/reassembly: `currentPacket` and `lastPacket`. The Reply body is composed by the result type and its value for both formats: Internet and WSN (see Figure 9(b)).

Besides saving energy by its reduced size, the sensor message format also concerns about sensor limited processing as it is already deployed as a byte array, avoiding the need for a Marshaller implementation.

The WISeMid Naming Service, the Trader and Internet services use the Internet format ($WIOP_i$), while the sensor services use the WSN format ($WIOP_s$). Only SAGe handles both formats.

4.3.2. Naming service

The WISeMid Naming Service implements the Naming discovery procedure defined by Smart in Section 3.3.3.1. It stores the references of services executing in the Internet and WSN in such a way that a service may only be accessed/used after being registered in the Naming Service.

The Naming Service's interface includes five operations: Bind, to register a service by its name, associating it with its reference; Lookup, to return the reference associated to a service name; Rebind, to change the reference that is associated with a service name; Unbind, to unregister a service name; and List, to list all registered services. The service reference includes the service

ID, endianness and the Internet end point of the service, which consists of its IP address and port number. Note that, since the Naming service runs in the Internet, the service references handled by it have no information about WSN end points. Nevertheless, WSN services may also be registered using the Internet end point of SAGe, which handles another type of service reference to store WSN end point information, as described next. It is worth mentioning that WISeMid Naming Service has a flat architecture, with a global name space.

4.3.3. Trading service

WISeMid's Trader implements the Trading service, another discovery procedure described by Smart (see Section 3.3.3.2). It has a graphical user interface (GUI) that allows users to define service types and property types, as well as export and query service offers.

The service type comprises a name, a reference for the service interface, a parent and a list of property types (those last two are optional attributes). When defined, each property type comprises a name, a property value type (e.g., integer, string, float, etc.) and a mode (which defines whether a property is mandatory and/or readonly) (see Section 3.2.1).

A service offer is the information that is registered in the Trader about a service and the act of registering that information in the Trader is called exporting a service offer. As stated in Section 3.3.3.2, a service offer consists of the service type name, a reference to the interface that provides the service, and zero or more property values for the service, where those properties correspond to the ones that are listed in the associated service type and specifying a value to them depends on the mode associated to each one of them. A value must be provided for every mandatory property (both `Mandatory` and `MandatoryReadOnly` modes) of the service type. On the other hand, providing a value to the other properties is optional, and properties that have not been specified in the related service type may be added in the service offer (with values being provided to them). WISeMid implements all those characteristics and adds a service offer identification number to uniquely identify a service offer.

4.3.4. Energy-saving methods

Since energy is a scarce resource in WSNs, performing energy-saving methods is essential to extend the sensor nodes lifetime and thus the WSN lifetime. This section presents some energy-saving approaches that have been implemented in WISeMid [9].

4.3.4.1. Aggregation

Besides this in-network data aggregation, which is used by most middleware, WISeMid implements an aggregation service which aggregates the last `n` data sensed by a node. The effect of this service is to aggregate results of services provided by a sensor node, sending only one (1) reply message instead of `n`. In addition to avoiding the transmission of `n-1` reply messages, this procedure also eliminates the need of sending `n-1` requests. The service user may send just one request to receive the same `n` values but in an aggregated form.

4.3.4.2. Reply storage timeout

Considering that some physical aspects sensed by a sensor node, such as temperature, do not present a great variability in terms of second/minute time scale, this approach avoids sending equivalent Request messages (that is, messages asking for the same service with the same parameters) during a short period of time, as the returned data are likely to be the same. Hence, SAGe groups equivalent Request messages and, for a configurable period of time, only one Request is sent to the sensor service provider, and the received Reply message is stored and forwarded as an answer to all the equivalent Request messages that arrive during that period. For the cases where the sensed value changes very often, this procedure may be turned off by setting to null (i.e., 0 seconds) the Reply message storage timeout.

4.3.4.3. Automatic type conversion

Since the transmission/reception of data is very energy consuming, the more data is sent/received by a sensor, the more energy is spent. With that in mind, SAGe performs an additional step when converting an Internet Request message into a sensor Request message. For each argument in the Request body, it tries to fit the argument value in a smaller type (that is, a compatible type that uses less bytes). For instance, if the argument is a `long` (an integer of 8 bytes) but its value is '123', it can be stored into a `byte` (an integer of 1 byte). Thus SAGe converts the argument from a `long` into a `byte` and adds only 1 byte to the $WIOP_s$ instead of the original 8 bytes, avoiding the transmission of 7 bytes. The same step is performed for the result value of $WIOP_i$ Reply messages when converting them into $WIOP_s$ Reply messages.

4.3.4.4. Asynchronous invocation patterns

Initially, the WISeMid services communicated synchronously. However, synchronous invocation blocks the service user (client) until the result returned from the service provider (server) is received. Thus, when a service user is running in a sensor node, it keeps consuming energy while waiting for the answer without executing any task. For that reason, we extended WISeMid with asynchronous invocation patterns, which enable the client to resume its work immediately after a remote invocation is sent. Four asynchronous invocation patterns, which are presented in [15], have been implemented in WISeMid, namely: Fire and Forget, Sync with Server, Poll Object and Result Callback. The first two patterns are used for one-way operations, while the last ones are used for request-response operations. Details on WISeMid implementation of those patterns can be found in [9].

5. SAGe

As stated previously, SAGe is an important WISeMid service which performs different tasks to allow the integration of the Internet and WSNs.

SAGe may be considered an advanced version of SerialForwarder [24], a TinyOS application which allows multiple clients to communicate with a mote working as a base station. SAGe extends SerialForwarder functionalities by performing some data processing before forwarding the packets from the WSN (more specifically, the sink node) to the connected clients and vice-versa. Running in the Internet host connected to the WSN sink node via a serial port (see Figure 7(b)), SAGe's main function is to act as a service proxy between both networks by enabling the communication between services running on Internet hosts and WSN nodes in a transparent way. Furthermore, SAGe also performs some tasks concerning the limited resources of sensor nodes, such as reducing the messages size and avoiding sending unnecessary messages to the WSN. The rest of this section describes SAGe architecture and functions.

5.1. SAGe architecture

The SAGe architecture, which is illustrated in Figure 10, comprises three components:

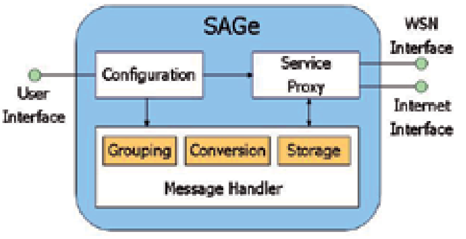


Figure 10. SAGe architecture

- **Configuration:** provides a user interface (UI) that enables to configure some SAGe attributes, such as: the port number used by Internet clients to connect to SAGe, and the period of time that a sensor Reply message should be considered an up-to-date information (to refrain SAGe from sending unnecessary Request messages to WSN).
- **Service Proxy:** provides one interface per network (WSN and Internet) through which SAGe receives/sends WIOP messages to provide communication between the services.
- **Message Handling:** handles WIOP messages and is divided into three subcomponents:
  - **Grouping:** groups equivalent Internet Request messages during a given period of time (set by Configuration) to avoid WSN energy waste;
  - **Conversion:** performs the conversion between Internet and WSN WIOP message formats (namely, $WIOP_i$ and $WIOP_S$);
  - **Storage:** stores the messages received from both networks during a given period of time (set by Configuration), in order to enable message grouping.

Those (sub)components cooperate to perform SAGe functions, which are described next.

5.2. SAGe functions

As mentioned above, the main functions of SAGe are acting as a service proxy and participating in the energy-saving mechanisms.

In order to enable the interaction of services of both Internet and WSN in a transparent way while acting as a service proxy, SAGe has an important role in three specific tasks: the binding of a WSN service in the Naming Service, the invocation of a WSN service (provider) by an Internet service (user), and the invocation of an Internet service (provider) by a WSN service (user). Those three tasks as well as SAGe's participation the methods to conserve the WSN nodes energy are described as follows.

5.2.1. Binding of a WSN service

Once a WSN service starts, it sends a message invoking the Bind operation of the Naming Service. When SAGe receives that message, it creates a `ServiceReference` to the WSN service including the SAGe's IP address and port, and then sends a bind request to the Naming Service, registering the WSN service as a SAGe service. It also keeps the created reference cached as a `SageServiceReference`, which assigns the `ServiceReference` to the node ID of the sensor providing the service. In such manner, SAGe knows which sensor node a Request message to a WSN service must be forwarded to.

5.2.2. Invocation of a WSN service

When the SAGe receives a Request message from the Internet invoking a WSN service, it converts the message to a $WIOP_s$ Request and sends it to the WSN using the `SageServiceReference` that has been cached when the WSN service was bound. Once the Reply message from the sensor service provider is received, SAGe converts it into a $WIOP_i$ Reply message and forwards it to the Internet service user. If no `SageServiceReference` is found, SAGe does not forward the request to the WSN since no sensor provides this service. Instead, it sends a Reply message reporting an error to the Internet host that requested the service.

5.2.3. Invocation of an Internet service

When a sensor node service performs a lookup for an Internet service, SAGe checks if this service is already known, i.e., if its reference is cached. If the service is unknown, SAGe converts and forwards the lookup request to the WISeMid Naming Service. When it receives the $WIOP_i$ Reply, it stores the returned `ServiceReference` and sends the service ID to the sensor node service. Using the received service ID, the WSN service invokes the Internet service operation. When SAGe receives the sensor Request message, it uses the cached `ServiceReference` to invoke the requested operation and, once the Reply message arrives, SAGe converts and forwards it to the sensor node service user.

5.2.4. Saving WSN limited resources

As exposed in Section 4.3.4, WISeMid implements some energy-saving mechanisms. SAGe has an essential role in some of these mechanisms, as explained below.

In the Reply Storage Timeout mechanism, SAGe is in charge of everything: allowing the configuration of the period of time the Reply message will be considered up-to-date; recognizing and grouping equivalent Request messages; storing and forwarding the Reply message for every Request message that arrives in the configured period of time.

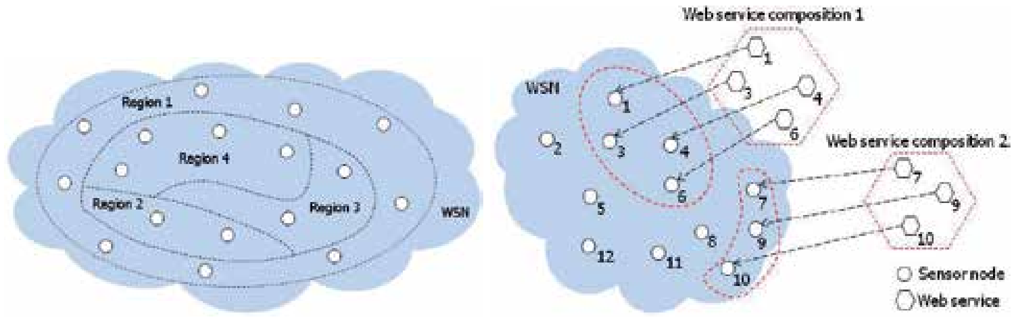
The Automatic Type Conversion is also performed by SAGe. When converting a $WIOP_i$ Request message into a $WIOP_s$ one, it tries to fit each argument value in the Request body in a compatible type that uses less bytes. The same procedure is performed for the result value of $WIOP_i$ Reply messages when converting them into $WIOP_s$ Reply messages.

Besides its active participation in those mechanisms, SAGe also performs small actions that help saving the sensor nodes energy. An example has already been described in Section 5.2.2: SAGe does not forward to WSN any Internet Request which asks for a sensor service that has not been registered. It would be useless and energy wasting since no sensor announced that service. Another example occurs when a sensor requests an Internet service (Section 5.2.3). It consists in not giving up at the first unsuccessful attempt to connect to the Internet service provider. Considering that it may be a sporadic problem, SAGe tries to connect to the server a configurable number of times before returning an error to the sensor node. This procedure aims to refrain the sensor from sending another Request in case the answer is fundamental for its application. Details of SAGe implementation are given in [10].

6. Clever

Several applications require particular attention to specific parts (regions) of a WSN. For instance, monitoring applications like a security surveillance may define the border region as a critical one to be monitored, whereas forest fire detecting applications may define clearings that resulted from previous fires as the critical regions to be monitored (since they tend to be more propitious for new fires due to their dry vegetation and soil) [25]. For this purpose, the WSN may be divided into logical regions, as illustrated in Figure 11(a).

Considering the given examples, Region 1 would be defined as the critical area for the security surveillance application and Region 4 would represent a clearing and be the critical area of the forest fire detection application.



(a) Logical regions in a WSN

(b) Service composition defining regions in a WSN

Figure 11. Defining logical regions in a WSN

With that in mind, Clever is added to the framework as a tool that allows defining these logical regions by composing Web services available in the Internet. Although the notion of logical regions have already been presented in the literature, to the best of our knowledge this is the first approach that uses web service composition to define logical regions in WSNs.

It is worth mentioning that regions intersection may occur since a single web service (which represents a sensor node) may participate in more than one composition. The next section describes the composition approach used by Clever.

6.1. Composition approach

Besides WISeMid services (that is, Internet's or WSN's services which use WIOP messages to interact with each other), Web services are also supported by WISeMid. In other words, WISeMid enables transparent communication among WISeMid services and Web services. That is possible by using two different kinds of proxy: one that plays the role of a Web service, allowing other Web services to access the WISeMid service it represents, and one that acts as a WISeMid service, to enable the communication between other WISeMid services and the Web service it is related to. Both kinds of proxies are automatically generated by ProxiesGen from WIDL and WSDL files and have the same operations that the services they represent.

Considering that, each WSN node may provide a (WISeMid) service, which may be represented by a Web service that acts as a proxy. Therefore, composing Web services which represent services that are running on sensor nodes can result in the creation of a logical region in the WSN. Figure 11(b) illustrates two Web service compositions that define two logical regions in a WSN. On both compositions, when an operation (e.g., getTemperature) belonging to the interface of the Web service composition is invoked, the invocation is passed to all sensor nodes belonging to the WSN's logical region and represented by the composition.

As explained above, the Web services depicted in Figure 11(b) are actually proxies to WISeMid services running in the sensor nodes. Thus, the Web services that constitute the composition (which is itself a Web service) do not access directly the sensor nodes. Rather, they convert and forward the received requests to the associated WISeMid service provided by the nodes.

When many sensors provide the same (WISeMid) service, only one Web service is used as a proxy to them. That happens because the Web service proxy is automatically generated by the ProxiesGen based on the service's WIDL, which is identical for all sensors providing the same service. Hence, to create a logical region with sensors that provide the same service, a user must add that service `n` times, where `n` is the number of sensors that will comprise the region.

To identify to which sensor a request needs to be sent, a parameter is set in the URL of the Web service. For example, if the URL of a temperature Web service is

`http://localhost/jaxws-Temperature/Temperature`

then the resulting parameterized URL is

`http://localhost/jaxws-Temperature/Temperature?id=1` or

`http://localhost/jaxws-Temperature/Temperature?id=2`

where the id parameter is the identifier of the WSN node that provides the service.

The process of defining a WSN logical region by creating a Web service composition comprises three steps: (1) to choose the Web services to be used in the composition; (2) to set the order and the parameters for each Web service invocation; and (3) to set the URL parameter (node id) of each web service invocation. The first two steps are commonly used in Web service composition and are typically performed using BPEL and a modeling and verification tool that supports it. The third step requires modifying the WSDL file that results from the composition to add a partner link for each sensor that is supposed to compose the logical region, as explained in the following.

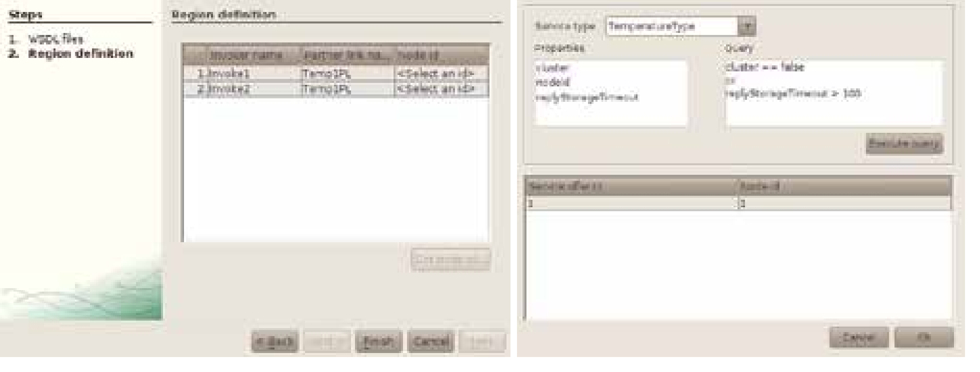
The BPEL partner links are the Web services used in the composition. Their URLs are static by default and obtained from the respective WSDL file. However, in the case that a Web service represents more than one WSN service, it is necessary to change these URLs updating dynamically the sensor identification number parameter, as there will be only one web service for them. This process is called Dynamic Addressing and an example of code is shown below:

```
<assign name="SetupPartnerlink"><copy>
 <from><literal><EndpointReference xmlns="http://schemas.xmlsoap.org/ws/2004/08/addressing"
      xmlns:tns="http://stub.mid/">
  <Address>http://localhost:8081/jaxws-Temp/Temp?id=6</Address>
  <ServiceName PortName="TempServicePort">tns:TempServiceWS</ServiceName>
 </EndpointReference></literal></from>
 <to variable="partnerReference"></to>
</copy></assign>
<assign name="AssignPL"><copy>
  <from>bpws:doXslTransform('urn:stylesheets:wrap2serviceref.xsl', $partnerReference)</from>
  <to partnerLink="Temp1PL"/>
</copy></assign>
```

```
12:<invoke name="Invoke1" partnerLink="Temp1PL" operation="getTemp" xmlns:tns="http://stub.mid/"
         portType="tns:TempService" inputVariable="GetTempIn" outputVariable="GetTempOut"/>
```

First, the new URL is assigned to an EndPointReference variable (1-7). Then, this variable is copied to a partner link (8-11). Finally, the invocation to the updated partner link is done (12).

The Clever tool is responsible for automating the third step by allowing the user to choose the nodes that should be part of the region and then automatically creating a partner link for each selected sensor node. Clever is implemented as a NetBeans plug-in, being accessed through a button. When that button is pressed, a wizard is opened to guide the user in choosing the nodes that will join the logical region. First, the Web service that represents the desired WSN service must be selected from a list of the services that have been added to the composition. Actually, the listed names are the partner link names of the invokes contained in the composition, with their respective WSDL file location. Once the Web service (partner link) is selected, the next window shows all the invokes to that service, as illustrated in Figure 12(a). For each invoke, a node ID must be set. To choose the node of interest, the user may search the WSN for all nodes that satisfy certain constraints, which are represented by property values of the offered service. For that purpose, Clever provides an interface that communicates with the Trader, as depicted in Figure 12(b). Using this window, the user is able to perform a query in the Trader, receiving a list of the nodes that satisfy that query to choose one of them.



(a) Step 2 for defining a logical region: selecting the nodes

(b) Finding and setting the node ID

Figure 12. Clever interface

When every invoke has a node associated, Clever automatically updates the BPEL file (of the Web service composition) to change the URL parameter before each partner link invocation. After that, the new (composite) Web service is ready and the related WSN logical region is created. Once the Web service composition is designed and deployed, a new Web service is created representing that composition. An invocation to this service corresponds to invoking each Web service involved in the composition and those Web services in turn request the related WSN service provided by each selected sensor node.

It is worth mentioning that, although the approach focuses on creating compositions with WSN Web services (more specifically, Web services representing WSN services) in order to form the logical regions in the WSN, it is possible to include Internet Web services in the composition, creating heterogeneous compositions. This possibility raises the level of integration since the (composite) service itself is distributed among nodes of both networks.

7. Energy consumption evaluation

This section presents results of experiments that analyze how WISeMid (including SAGe and the energy-saving mechanisms) and Clever affects the energy consumption in a sensor node.

For all scenarios, we use one Iris mote connected to a MTS400 basic environment sensor board, running the application defined in the studied scenario; and one Iris mote connected to a MIB520 USB programming board, working as a base station (BS), i.e., the sink node. The BS is connected to an Internet host that runs SAGe. Also, two other services run on an Internet host: the Naming service and the service/application under evaluation.

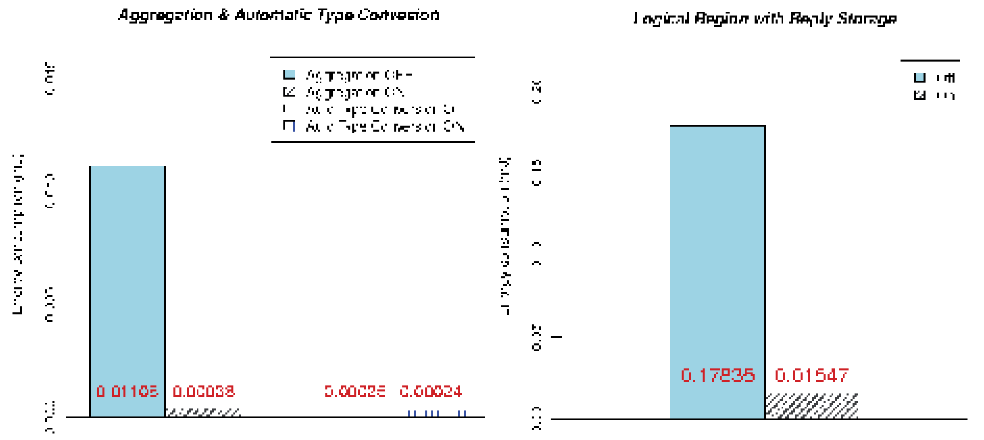
To estimate the energy consumption of the sensor node, an oscilloscope (Agilent DSO03202A) has been used. A PC is connected to the oscilloscope that captures the code snippet execution start and end times by monitoring a led of the sensor, which is turned on/off to signalize execution start/end. The PC runs a tool called AMALGHMA [26], which is responsible for calculating the energy consumption. In order to make the results more reliable, all values presented here are actually a mean value of 100 executions of the code in study.

Previous evaluation results have shown that WISeMid infrastructure has an impact over the energy consumption. Comparing an application that uses WISeMid infrastructure to a similar application that uses only TinyOS, the results shows that the service abstraction provided by WISeMid brings an energy consumption increase of 16.11% compared to the TinyOS application version (see [8] for details). Although it is not a negligible increase, the facilities offered by the WISeMid service abstraction as well as the energy saving brought by the energy aware mechanisms, which are presented next, compensate that.

Aggregation: To analyze the energy-saving offered by the Aggregation service, a scenario composed by an Internet service user requesting a WSN service that provides the sensed temperature is studied. When the Aggregation service is off, the Internet service sends 30 requests in a row; when it is on, only one request is sent by the Internet service user, and the WSN service provider returns one value which corresponds to the aggregation (average) of the last 30 sensed values. As Figure 13(a) shows, the Aggregation service saves 96.57% of energy. That significative rate is due to avoiding the transmission of 58 messages (29 requests and 29 replies), which would be necessary if the Aggregation service was not available.

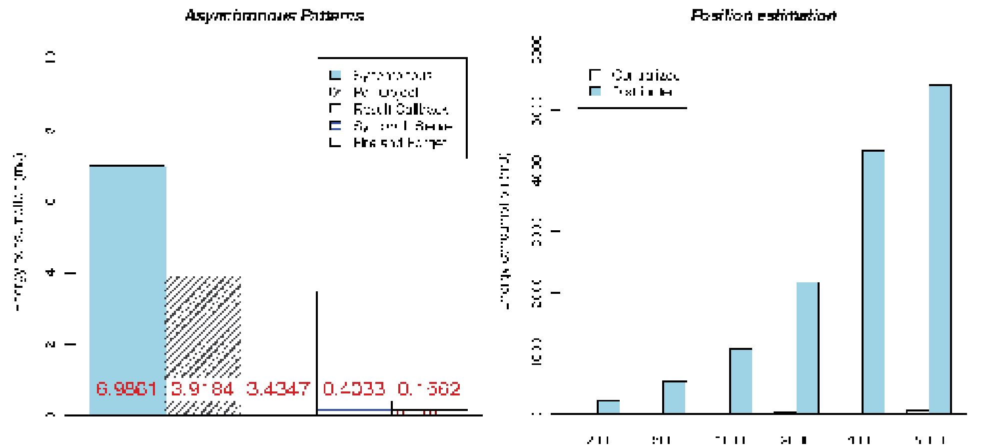
Automatic Type Conversion: The evaluation scenario for the Automatic Type Conversion comprises a simple WSN service, which has an operation that sets an attribute of type `long` (an 8-byte integer) and returns an acknowledgement, and an Internet service user that requests this operation with value "1" (one). As this value fits in type `byte`, when the Automatic Type Conversion feature is on, SAGe converts the parameter type from `long` to `byte` and only one byte is used to transmit the value "1", instead of 8 bytes. The results are illustrated in Figure 13(a) and show that, for the described scenario, the energy-saving gain is 4% when the Automatic Type Conversion is performed. To confirm whether that decrease in energy consumption is statistically significant, we performed a Paired t-test with those measurements. The resulting p-value is less than 4.352e-05, indicating the means are really different. Although 4% is not a very expressive gain, it is still an important gain as every little bit of energy that is saved in a WSN contributes to preserve and extend its lifetime.

Reply Storage Timeout in Logical Regions: The next scenario involves the use of a WSN logical region composed by two sensor nodes providing a Temperature service, which may also be invoked as a web service called TemperatureWS. As explained previously, only one proxy



(a) Aggregation service and Automatic Type Conversion

(b) Reply Storage Timeout used in Logical Regions



(c) Asynchronous Patterns instead of Synchronous communication

(d) Sensor position estimation method being performed by the WSN or by the Internet

Figure 13. Energy consumption of energy-saving mechanisms and real application

represents all sensors that provide the same service, hence TemperatureWS represents both sensors' service. A Web service composition (WSComposition) with the two Temperature services (actually, the TemperatureWS proxy) is created using Clever for selecting the nodes of interest. Finally, a service that runs on the Internet (TempMon) and monitors the temperature in that area calls WSComposition, which actually calls TemperatureWS twice in a row, once for each sensor node. The proxy calls the Temperature service in the sensors and then forwards their replies to the WSComposition, which calculates the mean value of the received results and returns it to the TempMon service.

To evaluate this logical region composition with the Reply Storage feature, the TempMon service requests the Temperature service composition (WSComposition) 25 consecutive times with random interval between messages, and a timeout of 4s. Figure 13(b) presents the results, which show that the Reply Storage feature saves 91.28% of energy in the studied scenario, which is a very significant energy-saving rate.

Asynchronous invocation patterns: The asynchronous patterns are evaluated in groups according to their type of operations: one-way or request-response. The one-way scenario is composed by a Logging service provided by an Internet host that is used by a WSN service to record log messages. The patterns evaluated here are Fire and Forget and Sync with Server.

The request-response scenario consists of the first distributed application described in Section 2, with a Surveillance service running in a sensor node, which captures images of any moving target and sends them to a Recognition service running in the Internet.

The results of both scenarios are presented in Figure 13(c). For one-way operations, a sensor saves 61.27% of energy when Fire and Forget is adopted instead of Sync with Server. That significant difference was expected since the former ends the process the moment the invocation is sent, whereas the latter involves the tasks performed in SAGe (such as message translation) plus the round-trip delay time from SAGe to the server. For response-reply operations, comparing to synchronous communication, the Poll Object pattern saves 43.91% of energy whilst the Result Callback obtains 50.83% of energy-saving. Those results confirm that the asynchronous invocation patterns can be adopted to extend the WSN lifetime.

Requesting Internet Services: One of WISeMid's main characteristics is that it allows sensor nodes to be not only service providers, but also service users. There are several situations in which asking for a node outside the network to perform a task is more suitable than performing it in-network. It may be even less costly in terms of energy consumption to request the service outside the WSN. Next scenario evaluates an environment monitoring application that shows this strategy. This application is composed by a Monitoring service that runs in a mobile sensor node and therefore needs to know the node position before reporting measured data (see [27] for examples of real applications with mobile sensor nodes). The Monitoring service may estimate the node position or it may request the position estimation from another service (called Positioning), which is provided by an Internet node. This scenario evaluates the energy consumption of the sensor node that provides the Monitoring service in both situations: performing and requesting the node position estimation.

To estimate the node position, we adopted an approach that has been developed with two versions: centralized and distributed. Both versions use measurements of distance between every node and its neighbors. In the centralized version, all sensor nodes send information about their neighbors to a central machine (outside WSN) with plenty of computation power where the nodes position are calculated and sent back to the network. In the distributed version, each node is responsible for calculating its position using information about its neighbors. Details of this localization approach, including implementation description and accuracy evaluation results, can be found in [28, 29].

Considering those information, we were able to estimate the energy consumption for both versions of this approach. The centralized version consumption comprises the energy that a node spends by requesting its location to a service that is provided by the Internet (Positioning service). It is worth mentioning that synchronous communication is being used, therefore this energy consumption includes the time the sensor keeps waiting for the reply. Moreover, to perform the localization algorithm, the Positioning service must have received the neighbor information of all WSN nodes. Hence, we have added to this calculation the energy consumption of all sensor nodes sending a message with their neighbors information.

For the distributed version, the values represent the estimated energy consumption of all messages that have to be exchanged for the position calculation. That value actually comprises the estimation of the number of messages that each node sends during all the localization algorithm execution, which involves several iterations in the optimization process [29] and different types of messages [28]. This estimated number of messages was used to measure the power a sensor node consumes to send all those messages, then this measured value was multiplied by the number of sensor nodes that compose the WSN.

Figure 13(d) shows the estimated energy consumption for both versions considering different sizes of network. As one can observe, in some situations, asking for a node outside the network to perform a task is more suitable than performing it in-network. For this specific application which uses this localization algorithm, the centralized version consumes approximately 98% less energy than the distributed version. That huge energy-saving rate is due to the great number of messages that have to be exchanged by the sensor nodes in the distributed version. In a network with 200 nodes, for instance, approximately 22,400 messages are transmitted.

8. Concluding remarks

This chapter introduced a framework for integrating WSNs and the Internet at service level, by allowing the interoperability of their services. The framework is composed by four components: Smart, a service model to describe services from both networks (Internet and WSNs); WISeMid, a middleware that enables the interoperability of WSN's and Internet's services; SAGe, an advanced gateway which is responsible for providing the location and access transparency for the services' communication; and Clever, a service composition tool which enables the creation of logical regions in the WSN by just composing Web services that are available on the Internet.

Some evaluation results concerning the energy consumption of those elements have been presented. Those results lead to the conclusion that by using some WISeMid's energy-saving mechanisms, a sensor node may save significant amount of energy. For instance, using the Aggregation service, a sensor node can save 96.57% of energy, whereas by using the Reply Storage Timeout feature in a logical region, a saving of 91.28% is obtained. Furthermore, a real application which requests an Internet service that estimates the sensor positioning instead of computing it in-network saved approximately 98% of energy, emphasizing the usefulness of the sensor nodes' capability of being service requestors.

In terms of future work, there are a number of possibilities to improve and extend the framework elements. The Smart model may be improved by detailing some elements, such as `MessageExchangePattern`'s `Publish/Subscribe`, and those elements may be implemented in WISeMid. Some features may also be added to SAGe, such as turning it into a distributed service, to refrain it from becoming a bottleneck in large-scale WSN. Also, some energy-saving mechanisms may be improved. For instance, the Reply Storage Timeout feature current implementation, which uses a global and fixed timeout value, would probably increase its energy gains by implementing an adaptive method to automatically update the timeout value per service, based on the variation history of the data sensed by the WSN nodes.

Author details

Jeisa Domingues
Federal Rural University of Pernambuco, Brazil

Antonio Damaso, Rilter Nascimento and Nelson Rosa
Federal University of Pernambuco, Brazil

9. References

[1] Rao S G., Zeldy S., Usman S., Mazlan A. A gateway solution for IPv6 wireless sensor networks. In: proceedings of the International Conference on Ultra Modern Telecommunications & Workshops (ICUMT '09), pp. 1-6; 2009.

[2] Smolderen K., De Cleyn P., Blondia C. Wireless Sensor Network Interconnection Protocol. In: proceedings of the IEEE Globecom 2010 Workshop on Heterogeneous, Multi-hop Wireless and Mobile Networks (HeterWMN 2010), pp. 164-168; 2010.

[3] Yoon I-S., Chung S-H., Jeong-Soo K. Implementation of Lightweight TCP/IP for Small, Wireless Embedded Systems. In: proceedings of the IEEE 23th International Conference on Advanced Information Networking and Applications (AINA), pp. 965-970; 2009.

[4] Bai J., Zang C., Wang T., Yu H. A Mobile Agents-Based Real-time Mechanism for Wireless Sensor Network Access on the Internet. In: proceedings of the 2006 IEEE International Conference on Information Acquisition, pp. 311-315; 2006.

[5] Samaras IK., Gialelis JV., Hassapis, GD. Integrating Wireless Sensor Networks into Enterprise Information Systems by Using Web Services. In: proceedings of the Third International Conference on Sensor Technologies and Applications (SENSORCOMM '09), pp. 580-587; 2009.

[6] Priyantha NB., Kansal A., Goraczko M., Zhao F. Tiny Web Services: Design and Implementation of Interoperable and Evolvable Sensor Networks. In: proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, pp. 253-266; 2008.

[7] Domingues J., Damaso A., Rosa N. Smart: Service Model for Integrating Wireless Sensor Networks and the Internet. In: proceedings of the 16th International Conference on Parallel and Distributed Systems (ICPADS), pp. 365-372; 2010.

[8] Domingues J., Damaso A., Rosa N. WISeMid: Middleware for Integrating Wireless Sensor Networks and the Internet. In: proceedings of the 10th IFIP international conference on Distributed Applications and Interoperable Systems (DAIS), pp. 70-83; 2010.

[9] Domingues J., Damaso A., Nascimento R., Rosa N. An Energy-aware Middleware for Integrating Wireless Sensor Networks and the Internet. International Journal of Distributed Sensor Networks 2011; Article ID 672313, 19 pages.

[10] Damaso A., Domingues J., Rosa, N. SAGe: Sensor Advanced Gateway for Integrating Wireless Sensor Networks and Internet. In: proceedings of the IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 698-703; 2009.

[11] Estefan J., Laskey K., McCabe F., Thornton, D. OASIS Reference Architecture for Service Oriented Architecture (SOA-RM), Version 1.0; 2008.

[12] Racz P., Stiller, B. A Service Model and Architecture in Support of IP Service Accounting. In: proceedings of the 10th IEEE/IFIP Network Operations and Management Symposium (NOMS 2006), pp. 1-12; 2006.

[13] Gracanin D., Eltoweissy M., Wadaa A., DaSilva L. A service-centric model for wireless sensor networks. IEEE Journal on Selected Areas in Communications 2005; 23 (6): 1159-1166.

[14] Rezgui A., Eltoweissy M. Service-oriented sensor-actuator networks: Promises, challenges, and the road ahead. Computer Communications 2007; 30 (13): 2627-2648.

[15] Völter M., Kircher M., Zdun U. Foundations of Enterprise, Internet, and Real-Time Distributed Object Middleware. New York: John Wiley & Sons, Inc.; 2004.

[16] Rubio B., Diaz M., Troya JM. Programming Approaches and Challenges for Wireless Sensor Networks. In: proceedings of the 2nd International Conference on Systems and Networks Communications, pp. 36-43; 2007.

[17] OMG. Trading Object Service Specification. OMG Technical Document Number 00-06-27; 2000.

[18] Benigni F., Brogi A., Fuentes T., Popescu R. D2.1: Service Model Description. SMEPP Consortium; 2008.

[19] Ran S. A model for web services discovery with QoS. SIGecom Exch. 2003; 4 (1): 1-10.

[20] Pokraev SV., Quartel, DAC., Steen MWA., Reichert MU. A Method for Formal Verification of Service Interoperability. In: proceedings of the IEEE International Conference on Web Services (ICWS'06), pp. 18-22; 2006.

[21] Sahai A., Singhal S., Machiraju V., Joshi R. Automated policy-based resource construction in utility computing environments. In: proceedings of the 2004 IEEE/IFIP Network Operations and Management Symposium (NOMS 2004), pp. 381-393; 2004.

[22] Deschrevel J-P. The ANSA Model for Trading and Federation. ANSA Architecture Report APM.1005.01; 1993.

[23] Schantz R., Schmidt D. Middleware for Distributed Systems: Evolving the Common Structure for Network-Centric Applications. In: Encyclopedia of Software Engineering, J. Marciniak, and G. Telecki, Eds. John Wiley & Sons, Inc., New York; 2002.

[24] Levis P., Gay D. TinyOS Programming. New York: Cambridge University Press; 2009.

[25] Carvalho Jr. JA., Veras CAG., Alvarado EC., Sandberg DV., Carvalho ER., Gielow R., Santos JC. Fire spread around a forest clearing site in the Brazilian Amazon Region. In: proceedings of the 2nd International Wildland Fire Ecology and Fire Management Congress; 2003.

[26] Tavares E., Silva B., Maciel P. An Environment for Measuring and Scheduling Time-Critical Embedded Systems with Energy Constraints. In: proceedings of the IEEE International Conference on Software Engineering and Formal Methods (SEFM), pp. 291-300; 2008.

[27] Romer K., Mattern F. The Design Space of Wireless Sensor Networks. IEEE Wireless Communications 2004: 11, 54-61.

[28] Marks M., Niewiadomska-Szynkiewicz E. Multiobjective Approach to Localization in Wireless Sensor Networks. Journal of Telecommunications and Information Technology 2009: 3, 59-67.

[29] Niewiadomska-Szynkiewicz E., Marks M. Optimization schemes for wireless sensor network localization. Int. J. Appl. Math. Comput. Sci. 2009: 19, 291-302.

Section 2

Data Management

Chapter 8

Data Reduction in Low Powered Wireless Sensor Networks

Qutub Ali Bakhtiar, Kia Makki and Niki Pissinou

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50178

1. Introduction

A Wireless sensor network (WSN) [1] is a special kind of a peer to peer network where the nodes communicate with the sink wirelessly to transmit the sensed information. In contemporary world, a WSN utilizes different technological advancements in low power communications and Very Large Scale Integration (VLSI) [16] to support functionalities of sensing, processing and communications. WSN's have penetrated in all walks of life ranging from health care, environmental monitoring to defense related applications. One of the major challenges faced by WSN's is that of energy conservation. Researchers all over the world have been trying their best to make sure that WSN efficiently make use of the energy resources and increase their life span. In this Chapter one such technique for energy conservation called 'Data Reduction' is discussed.

Data reduction can be defined as the process of conversion of all the information in a finite data set into fewer subsets and later on regenerating the entire set using the reduced subset. Data reduction is often the first step to tackle a data set because it facilitates in extracting its unique features. Typically, data reduction techniques are used for data mining large data warehouses.

In WSN, data reduction forces the sensor nodes to stop transmitting the data when it is confident about regenerating the future data at the sensor sink based on the existing, past and proximity observations thereby conserving the energy resources used for transmission of data.

A common way to facilitate data reduction in WSN is by deploying adaptation and prediction mechanism at the source and the destination so as to adapt to the changing pattern of the data and to predict it. The above mechanism is efficient in conserving the communication resources involved as it requires the source to relay only a subset of the actual data [17]. Moreover, since the radio transmission at the node consumes more amount of energy than any other operation at the node [1],[13] data reduction becomes an attractive option to conserve the limited energy resources of WSN.

Data reduction in WSN is a challenging process as data exists in the form of continuous stream (infinitely large data set) where the adaptation and prediction has to be performed online i.e. at a given instance of time not all the information is available for processing. Typically, the spatial and temporal relations among the data sources in WSN are exploited to achieve fair data reduction rates. Spatial and temporal characteristics of WSN's facilitate in adapting the environment and then predicting it thereby resulting in data reduction.

2. Data reduction systems in WSN

Data reduction systems have been deployed in many environments and scenarios to achieve conserve communication resources. Discrete fourier transforms (DFT) [21] is a classic technique used to perform data reduction for temporally related data streams. Based on DFT, researchers have more recently developed an advanced method Discrete Wavelet Transform (DWT) [5] to achieve data reduction. Also techniques such as, Singular Value Decomposition (SVD) [6] based on traditional Principal Components Analysis (PCA) [15] is an attractive data reduction technique because of its ability to provide optimal data reduction. Random projection of time series [4] is another technique which has exhibited great promise and demonstrated good results since it can provide approximate answers with guaranteed bounds of errors. Stated below are some of the prominent data reduction schemes presented in the literature for WSN.

2.1. Barbie-Q: A tiny-model query system (BBQ)

BBQ [9] is a data acquisition model for sensor networks that incorporates statistical models of real-world processes into a sensor network query processing architecture. The problem that BBQ addresses is that given a query and a model; identify a data acquisition plan for the sensor network to best refine the query answer. BBQ uses a specific model based on time-varying multivariate Gaussians in its architecture to compare it with the incoming data stream. If all of these probabilities generated meet or exceed user specified confidence threshold, then the requested readings are directly reported as the means in the probability density function. BBQ imparts confidence on the posterior density generated by time varying multivariate Gaussians and optimizes the expected benefit and cost of observing the attributes.

BBQ depends heavily on the prediction ability of the time varying multivariate Gaussians. In case of high nonlinearity among the sensed observations, the predictive ability of the multivariate Gaussians (markovian in case of dynamic environments) fails thereby responding the query incorrectly in spite of having a confidence in the model due to previous sensed observations.

On the guidelines of BBQ, the concept of multivariate Gaussians was exploited by Eiman et.al [10] in their design to propose a context (spatial and temporal) aware data cleaning mechanism. The proposed model by Eiman et. al. is more effective and efficient since it is contextually aware about the environment.

2.2. Probabilistic Adaptable Query system (PAQ)

Daniela et.al [25] proposed a general framework PAQ to efficiently answer queries at the sink based on a simple type of time-series model called an autoregressive model (AR) [21]. PAQ uses a combination of AR models to probabilistically answer queries. The model is used both globally, at the sink, to predict the readings of individual sensors, and locally, at each sensor, to detect when the sensor produces outlier readings or when the model ceases to properly fit the data. The sink maintains one AR model per geographic cluster and one cluster head in each cluster. The cluster head is responsible for communicating with the sink on behalf of its cluster. The cluster head's AR model, called the cluster model, is used to predict the values of all sensors in the cluster with an error of at most a fixed threshold over the member sensor's local models with the same confidence.

The sink maintains the coefficients associated with each of the leader's models and receives periodic readings from them. It also maintains a list of the current clusters. The cluster head's models and the cluster sets stored at the sink allow the sink to answer queries over all sensors using just the cluster models. To reduce communications, clusters are computed locally by the cluster heads and members of each cluster. Each cluster head is responsible for notifying the sink of changes in its model coefficients or in its cluster members, and for transmitting periodic readings to the sink.

Although the proposed mode achieves stable data reduction rates, one of the shortcomings of this approach is the communication of the AR parameters from the local cluster head to the sink. If the parameters do not reach the sink successfully or they get corrupted on the way then the AR models fails to synchronize at the node and the sink, ultimately causing improper prediction of the data.

2.3. Similarity-based Adaptive Framework (SAF)

Daniela et.al modified PAQ to incorporate the spatial attributes of sensor networks by using similarities among the node to propose SAF [24]. The premise behind SAF as similar to PAQ is to build local prediction models at each node, transmit them to the root of the network and use them to answer user queries. SAF also provides a mechanism to detect data similarities between nodes and organize nodes into clusters at the sink at no additional communication/complexity expense on the nodes which is absent in PAQ. This is achieved by exploiting properties of local time series models, and by means of utilizing data similarity between nodes that is based on the prediction values.

The prediction models in SAF are developed using lightweight linear time series models built by each node from a small number of readings (enabling models to be quickly re-learned) and stored at the sink. Sensor nodes and the sink communicate occasionally to exchange models or answer queries that require more accuracy than the stored models can provide. SAF attempts to group sensor nodes into clusters under dynamic conditions, which is a challenging problem as continuous monitoring and adaptation of cluster membership is required. The clustering algorithm used in SAF has another benefit that it does not require nodes in the same cluster to be geographically co-located, and does not require nodes to communicate at all with each other, thus making the clusters highly adaptable.

SAF is based on sound mathematical principles. However, it uses prediction variables to cluster the nodes according to the similarity based on prediction. Any misconduct exhibited by the temporal behaviour (a common feature in noisy wireless channels and hardware inefficient sensing devices) of the sensed phenomena introduces a certain amount of uncertainty while identifying the commonalities among the data, causing the entire model to fail. Also the exchange of time series parameters is initiated from the nodes and then passed on to the sink i.e the complexity due to the adaptation of the pattern is forced on to the nodes and moreover there is no mechanism present in the model that ensures that the time series prediction parameters have successfully reached the sink.

2.4. Adaptive Model Selection (AMS)

Santini et. al. proposed a lightweight, online algorithm AMS [22, 23] that allows sensor nodes to autonomously determine a statistically good performing model among a set of candidate models. AMS exploits the fact that, gathered sensor data is usually accepted to lie within a known error bound. A sensor node regularly collecting local measurements can fit a prediction model to the real data and communicate it to the sink, which can then use the model to compute estimates of future sensor readings. The rationale of this approach is to use complex prediction models only if they prove to be efficient both in terms of computation and achievable communication savings, and otherwise to rely on simpler models. AMS maintains a database of complex models at the sink and a mechanism to identify appropriate models depending on the computational ability and the requirement of the application.

AMS fails to demarcate the differences between adaptation and prediction mechanisms employed. The model uses a single complex algorithm to perform both the tasks although there is a significant difference between them. A faster adaptation algorithm and a robust prediction algorithm can always be used to overcome this disadvantage. The other drawback of AMS is that no clear information about the number and kind of complex models that are supposed to be stored has been specified. In reality it is not possible to store all the models that have the ability to perfectly emulate the environment that is being sensed.

2.5. Ken

Ken [7] is robust approximate technique that uses replicated dynamic probabilistic models to minimize communication from sensor nodes to the network's PC base station. Ken focuses on to intelligently exploit spatial correlations across sensor nodes without imposing undue sensor-to-sensor communication burdens to maintain the models. The basic premise behind Ken is simple: both source and sink maintain a dynamic probabilistic model of how data evolves, and these models are always kept in sync. The sink uses the data value(s) predicted by the model as the approximation to the true data, and the source, who knows the predicted value by virtue of running a copy of the model, makes sure that the predicted data values satisfy the required bounded-loss approximation guarantees, by communicating some information to the consumer as required. Ken uses disjoint clique approach to exploit the spatial correlations that exist among the nodes to achieve data reduction.

The main drawback of Ken is that complexity analysis has been completely ignored and is there no clear methodology to adaptively generate a probabilistic model. Ken justifies the claims made by using a Markovian approach where the information is incorporated in the model by conditioning (conditional probability). Ken achieves good results if the incoming data stream is linear in nature as it completely relies on conditional probability concept. However in reality the dynamics of the sensing information in its worst case are so strong that no fixed model is in a position to predict it. The non linearity exhibited by the sensed information causes the probability distribution to change and many times conditioning is not a reasonable approach to forecast the data successfully. In fact, incase of high nonlinearity there might be instances where conditioning further worsens the prediction ability of the model.

2.6. Prediction based Monitoring (PREMON)

PREMON [11] uses the coding concepts of MPEG that restores the quality of image using the partial image parameters. The work focuses on monitoring a sensor network in an energy efficient way by predicting all the readings which can be forecasted within a specified threshold of error. This work marked the beginning energy efficient monitoring of WSN by using prediction. The other novelty of this work is the usage of MPEG concepts used to increase the quality of ill formulated figure by utilizing efficient compression and coding techniques.

The entire process initiates by nodes sending the updates to the prediction model generator at the sink. The sink generates the model and passes the details of the model to the nodes which predict the sensed stream of data and halt the radio transmission saving the energy resources.

2.7. Dual Kalman Filter

Jain et.al proposed Dual Kalman Filter (DKF) [14] architecture as a general and adaptable solution to stream resource management problem. The architecture and the process is similar to PREMON as proposed by Goel et. al. the major difference between the usage of Kalman filter instead of MPEG concepts. The DKF model contains kalman filters at the sink synchronized with the kalman filter at each of the node. A major assumption made here is that the filter parameters are the same at the node and the sink at any given instance of time. The assumption is unreasonable for wireless and sensitive nodes as the communicating medium and radio resources are far from ideal.

Another drawback of the kalman filters is that the prior information about the input stream has to be provided as the input in order to train the filter. The size of training sequence is critical as in case of online data streaming systems the entire past data is not available at your disposal.

2.8. Approximate replication of data

Chris Olston in his PhD dissertation [17] proposed a data reduction method that operates on approximations of the sensor data distributions. The general approach presented by him achieves a substantial energy savings for the network as the data requirement is semi accurate in nature. The approximate mode of processing is extremely useful for two reasons. Firstly, it allows answering queries fast and cheaply, since all the nodes of the network relevant to the query need not be visited in order to get the answer. Secondly, it enables the execution of queries that would otherwise not be able to answer without consuming a lot more resources. The approximate requirement of the application also sheds off some load on the complexity and accuracy requirement of the prediction mechanism used.

Olston et. al. [18–20] proposed TRAPP (Trade off in replication precision and performance) architecture which maintains synchronous interval ranges at the sink and the node within the scope of the sensed observation. Each node is assigned a burden based on the data observation it senses. The burden represents the cost required to transmit the data to the sink. A similar approach has been used in WMA architecture for data cleaning where each node is assigned a weight based on its ability to provide a clean data.

Primarily, TRAPP deals with optimizing the precision and performance of the approximate data replication process between the node and the sink. The precision of the replicated data is inversely proportional to the interval where the value is detected to lie in. Although TRAPP has been developed for handling continuous queries its adaptive algorithm is also well equipped to handle the unprecedented one time aggregate queries that involves sum and the average functionalities. The adaptive algorithm reduces/increases the burden by increasing/decreasing the interval limit associated with each node and is said to converge in a steady state if the burden on all the nodes is uniform.

Although TRAPP provides huge energy savings for WSN, incase of one time queries that requires a response within a precision there is a high probability that it may fail to respond.

3. Data reduction in WSN using adaptive filters

Bakhtiar et. al. [2, 2] in their work demonstrated that data reduction in WSN can be achieved by deploying coordinating adaptive filters at the node and the sink to perform the operation of adaptation and prediction. The filters initially adapt to the varying nature of the sensed data until they converge. Convergence means that the error between the filter input and the filter output is within a specified threshold and the filters are ready to predict the future information.

After a successful convergence the node stops transmitting the data and the model moves to the prediction mode where the filters at both the ends begin to predict the future data. During the prediction mode the output of the filter at the sink is considered as the estimated sensed data, whereas the output of the filter at the node is compared with the sensed data to verify the prediction performance of the filter. The entire mechanism moves from the prediction phase to the adaptation phase if the error between the sensed data and the output of the filter at the node goes above a specified threshold.

In this Section, different adaptive filter algorithms presented in the literature [12] are used for Data Reduction in WSN. Primarily, the use of different classical estimation/prediction algorithms at the sink employed to reduce the set of transmitted data from the node is focussed. A special emphasis is also given on the initialization vector of the filters and a method to initialize the filters during the adaptation and the prediction phase is discussed. The presented model as stated in [2, 3] is tested using Least Mean Square LMS, Normalized Mean Square (NLMS), Recursive Least Squares (RLS) algorithms [12] at the sensor sink and the sensor node. Some modifications in the LMS algorithm employed during the prediction mode are also discussed and justification is made why the new modifications achieve better results than the traditional algorithm using appropriate simulations.

4. Model

Assume that there is a stream of data $\mathbf{x}(k)$ that is to be transmitted from a node to the sink

$$\mathbf{x}(k) = x(k)\ x(k-1)\ \ldots\ x(k-N) \tag{1}$$

Consider a N tap filter with the weights

$$\mathbf{W}(k) = \mathbf{w}_1(k)\ \mathbf{w}_2(k)\ \ldots\ \mathbf{w}_N(k) \tag{2}$$

where $\mathbf{w}_N(k)$ is the set of window taps at the k_{th} input sample. Let the desired signal be denoted by $d(k)$ which is the output value of the filter at $(k-1)_{th}$ instance during prediction and the sensed data during adaptation. The error in prediction is given by $e(k) = d(k) - y(k)$ where $y(k) = \mathbf{w}(k)\mathbf{x}^T(k)$ is the output of the filter.

Algorithm	Tap Update Function
LMS	$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{x}(k)$ $e(k) = [d(k) - y(k)]$
NLMS	$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu_n}{\lVert\mathbf{x}\rVert^2} e(k)\mathbf{x}(k)$ $e(k) = [d(k) - y(k)]$
RLS	$\mathbf{K}(k) = \frac{\rho^{-1}\mathbf{P}[k-1]\mathbf{x}(k)}{1+\rho^{-1}\mathbf{x}^H(k)\mathbf{P}[k-1]\mathbf{x}(k)}$ $\mathbf{P}[k] = \rho^{-1}\mathbf{P}[k-1] + \rho^{-1}\mathbf{K}(k)\mathbf{x}(k)$ $\mathbf{w}(k) = \mathbf{w}(k-1) + \mathbf{K}(k)e(k)$ $e(k) = [d(k) - y(k)]$

Table 1. Adaptive algorithms

The model discussed here is similar to that as presented in [3, 7, 23, 25] where coordinating filters are deployed at the sink and the node to adapt and predict the sensed information. The model is primarily divided into two modes, the adaptation mode and the prediction mode.

During the adaptation mode the filter uses the sensed information to adapt the sensing environment. It is during this mode that the sensed information is relayed to the sink and the filter at the node is idle. Once the filters are converged they begin to predict and if the difference between the predicted output and the sensed information remains below the threshold for N samples the model moves to the prediction mode. During the prediction mode the sensing nodes does not relay the sensed information to the sink, rather the estimated output at the sink is treated as the sensed data.

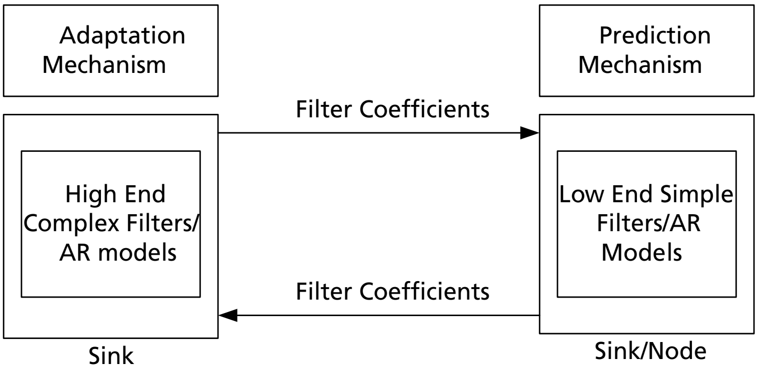


Figure 1. Flow diagram of the proposed model

In the methodology of data reduction presented in [3], the functionality of adaptation is restricted to sink only and the filter at the node remains idle until it receives the converged parameters of the filter from the sink. The authors then extend their previous work by incorporating high end complex filters at the sink during the adaptation mode [2]. The premise behind this approach is the ability of the sink to handle complex algorithms that offers fast and good convergence with an ease. Also, a simple prediction algorithm at the node to make sure that the energy resources at the node are utilized efficiently is also used in the above mentioned work.

The other major change of this work is the usage of filter coefficients during initialization of the mode. As shown in Figure 1, whenever there is a change in the mode of the model, the filter coefficients are exchanged. The filter parameters at the end of any mode are passed on to the filter of the other mode. These parameters serve as a initialization vector of the filter and assist in faster convergence.

As shown in Figure 2 the radio transmission of the node can be assumed to be controlled by the output of the prediction filter at the node. Each sensed sample is compared with the estimated output during the prediction mode and if the difference between the sensed data and the estimated output is below a specified threshold the switch is turned *off* i.e. the the node does not relay the sensed data to the sink. During this process the prediction filter at the sink estimates the sensed data and the approximate predicted output is used by the end application as the sensed information.

Incase, the difference between the predicted output and the sensed data goes above the threshold level, the node stops the prediction process and relays the sensed information to the sink. At this instance the switch is turned *on*. The sink uses the current filter parameters of the prediction filter and the sensed information to initialize the adaptation process and thus the entire model moves from the prediction mode to adaptation mode. The entire process repeats itself and the radio transmission at the node is accordingly controlled. The model is stable enough as any incapability of the prediction filter will result in the transmission of data from the node.

Another major change observed in this work is the usage of the prediction algorithm. During the prediction phase LMS algorithm is used only as it is the simplest and robust among

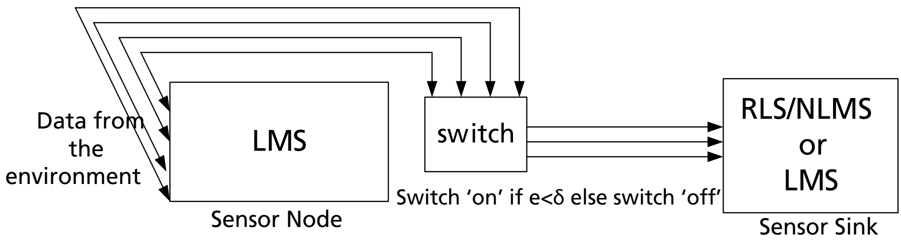


Figure 2. Transmission Control mechanism

the other liner prediction techniques. A more faster convergent algorithm can be used for prediction but faster convergent algorithms have more likelihood to diverge faster especially while performing prediction using the feedback mechanism. However, linear prediction techniques ideally forecast the nth value provided the filter has the prior information of $(n-1)$th values and if the prediction is extended to forecast the $(n+k)$th value during approximate replication using the feedback mechanism no significant change is observed in the output. The other problem that faced by LMS is that the step size parameter should have a common value during each prediction phase at the sink and the node.

In order to overcome the above mentioned drawbacks the authors modify the window tap update equation by removing the step size parameter from it. Intuitively, it can be said that the window tap update is representation of the input trend of data to the filter. It depends on the error between the output of the filter and the input at a particular instant which incase of data reduction is less than δ. The basic LMS equation is given by.

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{x}(k) \tag{3}$$

where μ is the step size parameter which can have values ranging from [12]

$$0 < \mu < \frac{2}{NS_{max}} \tag{4}$$

where S_{max} is the maximum value of the power spectral density of the tap inputs $\mathbf{x}(k)$ and N is the filter length. Assume that the value of the step size parameter be given by

$$\mu = \frac{1}{N(\mathbf{x}(k) * \mathbf{x}^T(k))} \tag{5}$$

During the prediction phase of the data reduction the value of $\mathbf{x}(k)$ changes minimally and the difference among the elements of the array $\mathbf{x}(k)$ is less than δ. Using the above facts and substituting the value of μ in Equation 3 the following tap update equation is obtained.

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{e(k)}{N\sum \mathbf{x}(k)} \tag{6}$$

The main objective of Equation 6 is to sub optimally predict the sensed information from the environment for a longer duration of time. The stability of the modified algorithm remains the same as that of LMS as the value of the step size parameter lies within the bounds of stability and moreover is use to educate the algorithm about the conditions of approximate replication.

5. Results

In order to prove the claims made, two simulation scenarios involving real time data sensed by two nodes from Intel lab [8] are performed on Matlab. Out of 54 nodes deployed in the Intel Lab, Node-4 and Node-11 is selectd for simulation. A total number of 43790 temperature samples are used in case of node 4 and 41833 in the case of Node-11. The threshold level is set to +/-0.5. The initial tap weights are assigned to be zeros. The length of the filter is chosen to be 3 (i.e. N=3). The value of μ for LMS is selected as 7.10^{-5} [23]. Let A_n be the number of times the model moves from the prediction mode to the adaptation mode. The model is tested by using the algorithms mentioned in Table 1.

Node 4 , 43790 Samples				Node 11 , 41833 Samples			
	LMS	NLMS	RLS		LMS	NLMS	RLS
Predicted Data	42228	42534	42904	Predicted Data	40495	41136	41178
Transmitted Data	1562	1256	856	Transmitted Data	1338	697	655
A_n	866	728	653	A_n	806	615	566

Table 2. Comparison of convergence rates for Node 4 and Node 11 using different algorithms when traditional LMS is used in the prediction phase

Initialization Method	LMS	NLMS	RLS
$W_{adapt}(n) = 0$	39179	40745	40942
$W_{adapt}(n) = W_{pred}(n)$	42228	42534	42904

Table 3. Comparison of different initialization methods on Node 4

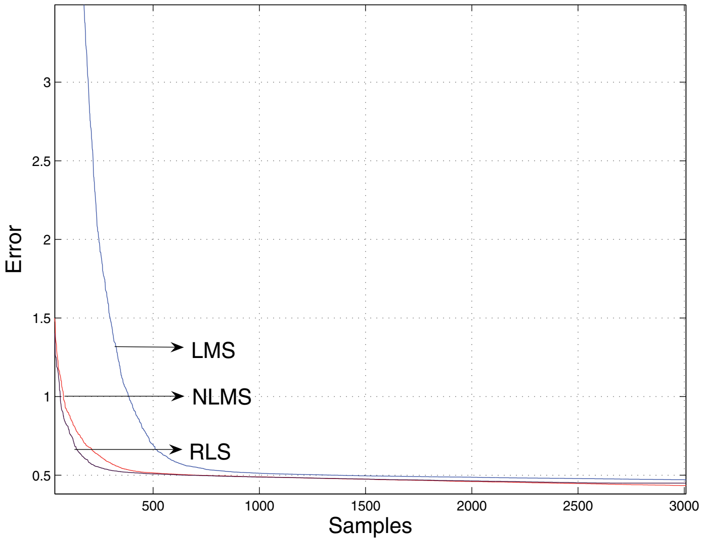


Figure 3. Convergence rates of the filters during adaptation mode at the sink using different algorithms for sensor node11

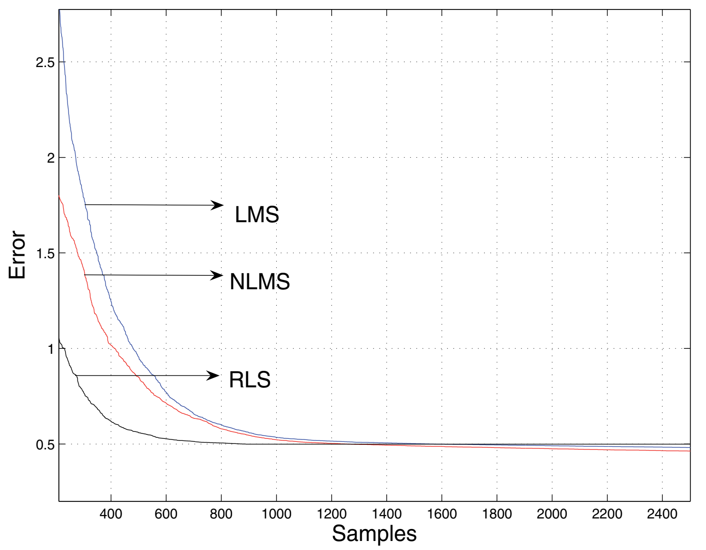


Figure 4. Convergence rates of the filters during adaptation mode at the sink using different algorithms for sensor node4

From Figure 3 and Figure 4 it is evident that the filter error (i.e the difference between the sensed data and filter output) of the sink converges faster when complex algorithms like NLMS, RLS are deployed. From Table 5 is is also observed that the value of A_n (the number of times the model shifts from the prediction mode to adaptation) is small when a more complex algorithm is used for adaptation. It is due to the fact that algorithms such has NLMS,RLS have better adaptation capabilities than LMS [12]. From Table 5 it can also be inferred that the number of samples required to approximately predict the remaining samples reduces when a better adaptation mechanism is deployed at the sink. As opposed to the previous model where only a simple adaptation and mechanism using LMS was used [23] due to limited node capabilities, a marked improvement in the energy savings is observed. This is due to the deployment of NLMS and RLS that achieves a better data reduction of 306 and 706 for node 4 and 641 and 683 samples respectively.

It can also be inferred from Table 5 that the initialization vector of the algorithm during adaptation has an impact on the rate of convergence. Whenever the process moves from the prediction mode to the adaptation mode, filter parameters of the prediction mechanism at the last instance to initialize the filter during the adaptation process are used. An improvement of 3049,1789,2412 while convergence is observed for LMS, NLMS and RLS algorithms when the sensed samples at node 4 were considered.

Figures 5 & 6 depict the window tap behaviour when traditional and modified LMS is used for prediction. It is observed that when traditional LMS is used a significant variation of window tap is achieved than when compared to the modified one. The traditional LMS does not exploit the conditions of data reduction and always tries to achieve an optimal solution.

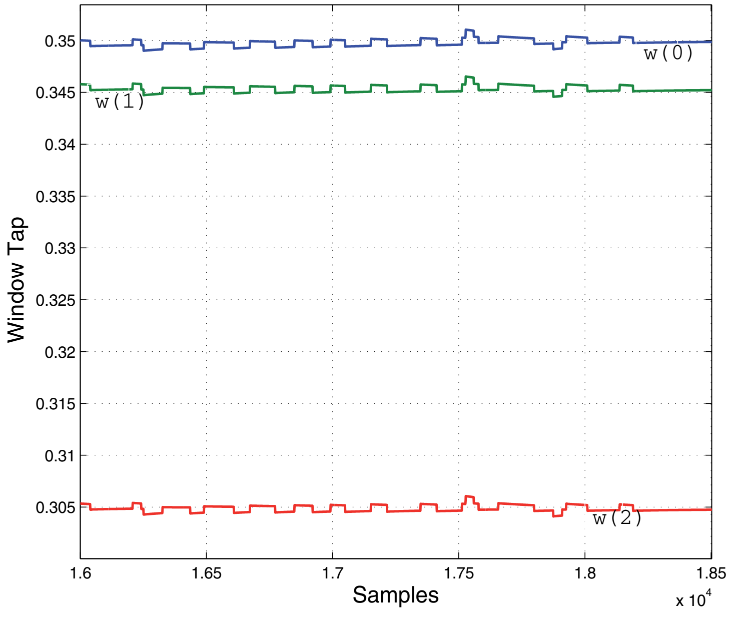


Figure 5. Amplitudes of the window taps when LMS is used for prediction

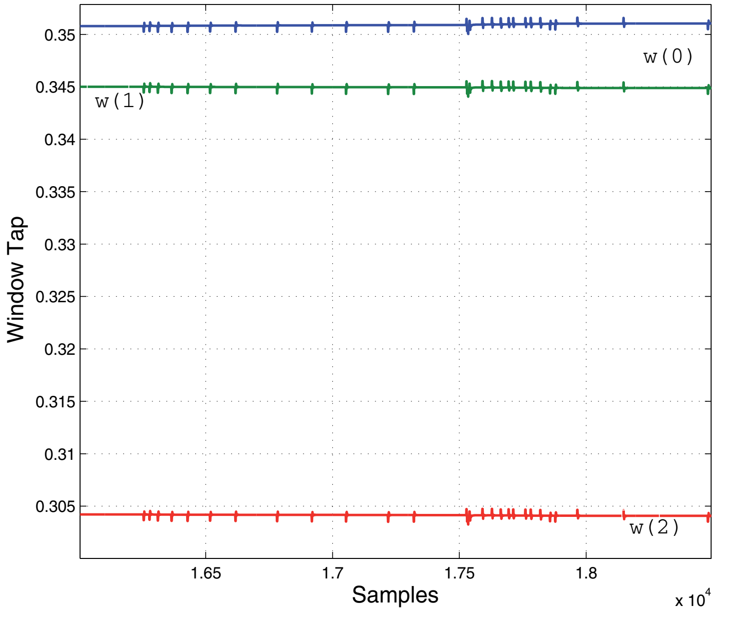


Figure 6. Amplitudes of the window taps when modified LMS is used for prediction

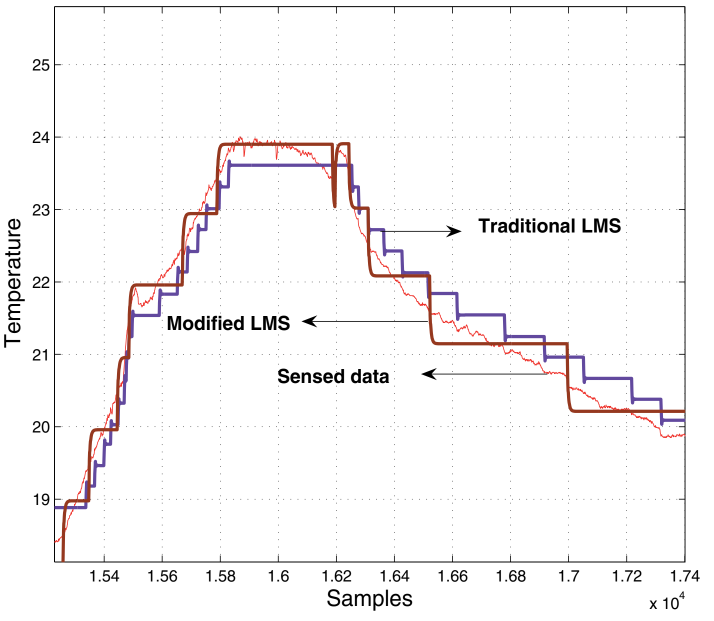


Figure 7. Comparison of modified LMS and traditional LMS when NLMS is used for adaptation at the sink

Node 4 , 43790 Samples				Node 11 , 41833 Samples			
	LMS	NLMS	RLS		LMS	NLMS	RLS
Predicted Data	42270	42998	43068	Predicted Data	40498	41270	41280
Transmitted Data	1520	792	722	Transmitted Data	1335	563	553
A_n	880	617	579	A_n	803	498	494

Table 4. Comparison of convergence rates for Node 4 and Node 11 using different algorithms with modified LMS in the prediction phase

Table 5 provides the predicted and the transmitted samples when modified LMS is used during the prediction phase. On comparison of Table 5 with Table 5 it can be concluded that whenever NLMS and RLS are used for adaptation, modified LMS should be used for prediction on the other hand when LMS is used for adaptation, traditional LMS should be used for prediction. The window taps exhibit a better shift from adaptation phase to prediction phase when LMS is used in both the phases.

6. Conclusion

This Chapter provides an overview of the data reduction systems that have been proposed in the last decade for Wireless Sensor Networks. In order to further elaborate the process of data reduction for WSN this Chapter delves into the dual filter model proposed by Bakhtiar et.al. [2, 2] that emphasizes on initialization of the filters and exchange of parameters between the filters. The model is mathematically elucidated by means of equations and tables. Also the environment is simulated and tested with real time data from Intel Labs [8]. The results obtained reaffirm the fact that data reduction obtains high communication resource savings and has a great impact in the extending the life span of the WSN. Although a significant amount of work is ongoing [26],[27],[28] on topics related to data reduction in WSN, there are open issues that can be addressed in the future, specifically when data reduction is achieved by means of adaptive filters. The issues are stated as follows

- **Synchronization** In data reduction models involving dual filters it is very important to maintain the synchronization of the filters at node and sink. By synchronization it is meant that at a given instance both the filters have to be in the same mode. In real time scenario synchronization adds an overhead to the communication process and any error or mistiming leads to the failure of the model.
- **Modeling the sensed data**: Prediction is usually carried out after modeling the data to some fixed form. There are many methods described in the literature that model the incoming data to some mathematical form. At the sink end this mathematical form can be used to predict the data.
- **Algorithms**: The present day algorithms do not fully exploit the threshold value restriction that is imposed at the sensing node. A new process which effectively exploits this condition can always be developed and deployed in the data reduction mechanism.
- **Spatial Considerations**: A filter which does the dual function of co-relation and fusion can be employed at the sensor sink by exploiting the spatial characteristics of the WSN. The correlation attribute will further enhance the data quality of the data output from the data reduction process.

Author details

Qutub Ali Bakhtiar and Kia Makki
Network Lab, Technological University of America, Coconut Creek, FL USA

Niki Pissinou
Information Technology and Telecommunications Institute, Florida International University, Miami, FL USA

Appendix: Derivation

The tradition LMS algorithm [12] can be stated as follows.

$$\mathrm{w}(k+1) = \mathrm{w}(k) + \mu e(k)\mathrm{x}(k) \tag{A.1}$$

where μ is the step size parameter which can have values ranging from [12]

$$0 < \mu < \frac{2}{NS_{max}} \tag{A.2}$$

where S_{max} is the maximum value of the power spectral density of the tap inputs $\mathrm{x}(k)$ and N is the filter length. Assume that the value of the step size parameter be given by

$$\mu = \frac{1}{N(\mathrm{x}(k) * \mathrm{x}^T(k))} \tag{A.3}$$

Let

$$\mathrm{x}(k) = \gamma \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{A.4}$$

The assumption of $\mathrm{x}(k)$ is valid in the case of prediction as data changes minimally. Substituting the value of and solving the equation

$$\mathrm{w}(k+1) = \mathrm{w}(k) \frac{e(k)}{N \sum \mathrm{x}(k)} \tag{A.5}$$

The above equation is devoid of the step size parameter μ which controls the speed and optimality of convergence of the algorithm.

7. References

[1] Akyildiz, I. F., Weilian, S., Sankarasubramaniam, Y. & Cayirci, E. [2002]. A survey on sensor networks, *IEEE Communication Magazine* 12: 102–114.

[2] Bakhtiar, Q. A., Pissinou, N. & Makki, K. [2008a]. Approximate replication of data using adaptive filters in wireless sensor networks, *International Symposium on Wireless Pervasive Computing*, Santorini, Greece.

[3] Bakhtiar, Q. A., Pissinou, N. & Makki, K. [2008b]. Estimated replication of data in wireless sensor networks, *International Conference on Communication Systems and Middleware (Comsware 08), Bangalore, India* .

[4] Bingham, E. & Mannila, H. [2001]. Random projection in dimensionality reduction: applications to image and text data, *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, New York, NY, USA, pp. 245–250.

[5] Burrus, C. S., Gopinath, R. A. & Guo, H. [1998]. *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice hall.

[6] Businger, P. A. & Golub, G. H. [1969]. Algorithm 358: singular value decomposition of a complex matrix [f1, 4, 5], *Commun. ACM* 12(10): 564–565.

[7] Chu, D., Deshpande, A., Hellerstein, J. M. & Hong, W. [2006]. Approximate data collection in sensor networks using probabilistic models, p. 48.

[8] Data, I. L. [n.d.]. *http://db.lcs.mit.edu/labdata/labdata.html*.

[9] Deshpande, A., Guestrin, J. H. C., Madden, S. & Hong, W. [2004]. Model-driven data acquisition in sensor networks., *30th VLDBŠ04, Toronto, Canada* .

[10] Elnahrawy, E. & Nath, B. [2003]. Cleaning and querying noisy sensors, *ACM International Workshop in Wireless Sensor Networks WSNAŠ03* .

[11] Goel, S. & Imielinski., T. [1994.]. Prediction based monitoring in sensor networks taking lessons from mpeg, *ACM Computer Communication Review* pp. 510–514,.

[12] Haykin, S. [2002]. *Adaptive Filter Theory*, 4th edn, Prentice Hall, Englewood Cliffs, NJ.

[13] Iyengar, S. & Brooks, R. (eds) [2005]. *Distributed Sensor Networks*, Chapman and Hall.

[14] Jain, A., Chang, E. & Wang, Y. [2004]. Adaptive stream resource management using kalman filters., *ACM SIGMOD/PODS Conference (SIGMOD Š04), Paris, France* .

[15] Jolliffe, I. [2002]. *Principal Component analysis*, Springer Berlin Heidelberg.

[16] Min, R., Bhardwaj, M., hwan Cho, S. & Sinha, E. S. A. [2001]. Low-power wireless sensor networks, *In VLSI Design*, pp. 205–210.

[17] Olston, C. [2003]. *Approximate Replication*, PhD thesis, Stanford University.

[18] Olston, C., Jiang, J. & Widom, J. [2003]. Adaptive filters for continuous queries over distributed data streams, *ACM SIGMOD, Intl Conference on Management of data (SIGMOD 03))* pp. 563–574.

[19] Olston, C., Loo, B. T. & Widom, J. [2001]. Adaptive precision setting for cached approximate values, *Proceedings of SIGMOD Conference* .

[20] Olston, C. & Widom, J. [2005]. Efficient monitoring and querying of distributed, dynamic data via approximate replication, *IEEE Data Engineering Bulletin* .

[21] Proakis, J. [2006]. *Digital Signal Processing*, Prentice Hall.

[22] Santini, S. [2006]. Poster abstract: Towards adaptive wireless sensor networks, *3rd European Workshop on Wireless Sensor Networks (EWSN 2006)* pp. 13–15.

[23] Santini, S. & Romer, K. [2006]. An adaptive strategy for quality-based data reduction in wireless sensor networks, *Proceedings of 3rd International Conference on Networked Sensing Systems(INSS 2006),Chicago* .

[24] Tulone, D. & Madden, S. [2006a]. An energy-efficient querying framework in sensor networks for detecting node similarities, *MSWiM '06: Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, ACM, New York, NY, USA, pp. 191–300.

[25] Tulone, D. & Madden, S. [2006b]. Paq: Time series forecasting for approximate query answering in sensor networks, *European Workshop on Wireless Sensor Networks EWSN* pp. 21–37.

[26] G. Patterson & M. Mehmet-Ali [2011] "Modeling of data reduction in wireless sensor networks," *Wireless Sensor Network*, vol. 3, no. 8, pp. 283–294, 2011.

[27] Rooshenas,A. , Rabiee,H.R. , Movaghar,A., & M.Y. Naderi,[2010] "Reducing the data transmission in wireless sensor networks using the principal component analysis," *In Proceeding of the Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP'10)*, (Brisbane, Australia), IEEExplore, 2010.

[28] E.Msechu & G.Giannakis, [2012 "Sensor-centric data reduction for estimation with wsns via censoring and quantization," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 400–414, 2012.

Chapter 9

An Algorithm for Denoising and Compression in Wireless Sensor Networks

Ehsan Sheybani

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48518

1. Introduction

Many wireless sensor network datasets suffer from the effects of acquisition noise, channel noise, fading, and fusion of different nodes with huge amounts of data. At the fusion center, where decisions relevant to these data are taken, any deviation from real values could affect the decisions made. We have developed computationally low power, low bandwidth, and low cost filters that will remove the noise and compress the data so that a decision can be made at the node level. This wavelet-based method is guaranteed to converge to a stationary point for both uncorrelated and correlated sensor data. Presented here is the theoretical background with examples showing the performance and merits of this novel approach compared to other alternatives.

Noise (from different sources), data dimension, and fading can have dramatic effects on the performance of wireless sensor networks and the decisions made at the fusion center. Any of these parameters alone or their combined result can affect the final outcome of a wireless sensor network. As such, total elimination of these parameters could also be damaging to the final outcome, as it may result in removing useful information that can benefit the decision making process. Several efforts have been made to find the optimal balance between which parameters, where, and how to remove them. For the most part, experts in the field agree that it is more beneficial to remove noise and/or compress data at the node level [Closas, P., 2007], [Yamamoto, H., 2005], [Son, S.-H., 2005]. This is mainly stressed so that the low power, low bandwidth, and low computational overhead of the wireless sensor network node constraints are met while fused datasets can still be used to make reliable decisions [Abdallah, A., 2006], [Schizas, I.D., 2006], [Pescosolido, L., 2008].

Digital signal processing algorithms, on the other hand, have long served to manipulate data to be a good fit for analysis and synthesis of any kind. For the wireless sensor networks a special wavelet-based approach has been considered to suppress the effect of noise and data order. One of the advantages of this approach is in that one algorithm serves to both reduce the data order and remove noise. The proposed technique uses the orthogonality properties of wavelets to decompose the dataset into spaces of coarse and detailed signals. With the filter banks being designed from special bases for this specific application, the output signal in this case would be components of the original signal represented at different time and frequency scales and translations. A detailed description of the techniques follows in the next section.

2. Wavelet-based transforms

Traditionally, Fourier transform (FT) has been applied to time-domain signals for signal processing tasks such as noise removal and order reduction. The shortcoming of the FT is in its dependence on time averaging over entire duration of the signal. Due to its short time span, analysis of wireless sensor network nodes requires resolution in particular time and frequency rather than frequency alone. Wavelets are the result of translation and scaling of a finite-length waveform known as mother wavelet. A wavelet divides a function into its frequency components such that its resolution matches the frequency scale and translation. To represent a signal in this fashion it would have to go through a wavelet transform. Application of the wavelet transform to a function results in a set of orthogonal basis functions which are the time-frequency components of the signal. Due to its resolution in both time and frequency wavelet transform is the best tool for detection and classification of signals that are non-stationary or have discontinuities and sharp peaks. Depending on whether a given function is analyzed in all scales and translations or a subset of them the continuous (CWT), discrete (DWT), or multi-resolution wavelet transform (MWT) can be applied.

An example of the generating function (mother wavelet) based on the Sinc function for the CWT is:

$$\psi(t) = 2Sinc(2t) - Sinc(t) = \frac{Sin(2\pi t) - Sin(\pi t)}{\pi t} \tag{1}$$

The subspaces of this function are generated by translation and scaling. For instance, the subspace of scale (dilation) a and translation (shift) b of the above function is:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{2}$$

When a function x is projected into this subspace, an integral would have to be evaluated to calculate the wavelet coefficients in that scale:

$$WT_{\psi}\{x\}(a,b) = \langle x, \psi_{a,b} \rangle = \int_R x(t)\overline{\psi_{a,b}(t)}dt \tag{3}$$

And therefore, the function x can be shown in term of its components:

$$x_a(t) = \int_R WT_\psi\{x\}(a,b).\psi_{a,b}(t)db. \quad (4)$$

Due to computational and time constraints it is impossible to analyze a function using all of its components. Therefore, usually a subset of the discrete coefficients is used to reconstruct the best approximation of the signal. This subset is generated from the discrete version of the generating function:

$$\psi_{m,n}(t) = a^{-m/2}\psi\left(a^{-m}t - nb\right). \quad (5)$$

Applying this subset to a function x with finite energy will result in DWT coefficients from which one can closely approximate (reconstruct) x using the coarse coefficients of this sequence:

$$x(t) = \sum_{m\in Z}\sum_{n\in Z}\left\langle x,\psi_{m,n}\right\rangle.\psi_{m,n}(t). \quad (6)$$

The MWT is obtained by picking a finite number of wavelet coefficients from a set of DWT coefficients. However, to avoid computational complexity, two generating functions are used to create the subspaces:

V_m Subspace:

$$\phi_{m,n}(t) = 2^{-m/2}\phi\left(2^{-m}t - n\right) \quad (7)$$

And W_m Subspace:

$$\psi_{m,n}(t) = 2^{-m/2}\psi\left(2^{-m}t - n\right). \quad (8)$$

From which the two (fast) wavelet transform pairs (MWT) can be generated:

$$\phi(t) = \sqrt{2}\sum_{n\in Z} h_n\phi(2t - n) \quad (9)$$

and

$$\psi(t) = \sqrt{2}\sum_{n\in Z} g_n\phi(2t - n) \quad (10)$$

In this paper the DWT has been used to suppress noise and reduce order of data in a wireless sensor network. Due to its ability to extract information in both time and frequency domain, DWT is considered a very powerful tool. The approach consists of decomposing the signal of interest into its detailed and smoothed components (high-and low-frequency). The detailed components of the signal at different levels of resolution localize the time and frequency of the event. Therefore, the DWT can extract the coarse features of the signal (compression) and filter out details at high frequency (noise). DWT has been successfully applied to system analysis for removal of noise and compression [Cohen, I., 1995], [Daubechies, I., 1992]. In this paper we present how DWT can be applied to detect and filter out noise and compress signals. A detailed discussion of theory and design methodology for the special-purpose filters for this application follows.

3. Theory of DWT-based filters for noise suppression and order reduction

DWT-based filters can be used to localize abrupt changes in signals in time and frequency. The invariance to shift in time (or space) in these filters makes them unsuitable for compression problems. Therefore, creative techniques have been implemented to cure this problem [Liang, J., 1996], [Cohen, I., 1995], [Daubechies, I., 1992], [Coifman, R., 1992], [Mallat, S., 1991], [Mallat, S., 1992]. These techniques range in their approach from calculating the wavelet transforms for all circular shifts and selecting the "best" one that minimizes a cost function [Liang, J., 1996], to using the entropy criterion [Coifman, R., 1992] and adaptively decomposing a signal in a tree structure so as to minimize the entropy of the representation. In this paper a new approach to cancellation of noise and compression of data has been proposed. The discrete Meyer adaptive wavelet (DMAW) is both translation- and scale-invariant and can represent a signal in a multi-scale format. While DMAW is not the best fit for entropy criterion, it is well suited for the proposed compression and cancellation purposes [Mallat, S., 1992].

The process to implement DMAW filters starts with discritizing the Meyer wavelets defined by wavelet and scaling functions as:

$$\phi(t) = \sqrt{2} \sum_{n \in Z} h_n \phi(2t - n) \tag{11}$$

and

$$\varphi(t) = \sqrt{2} \sum_{n \in Z} g_n \phi(2t - n) \tag{12}$$

The masks for these functions are obtained as:

$$\left\{ \phi(0), \phi(\frac{1}{2^m}), \ldots.., \phi(\frac{M-1}{2^m}) \right\} \tag{13}$$

and

$$\left\{ 0, 0, \ldots, 0, \varphi(0), \varphi(\frac{1}{\sigma}), \ldots.., \varphi(\frac{N}{\sigma}) \right\}. \tag{14}$$

As these two masks are convolved, the generating function (mother wavelet) mask can be obtained as:

$$F(\frac{k}{2^m}) \quad (-M \le k \le N). \tag{15}$$

Where for every integer k, integers $n_1^k, n_2^k, \ldots\ldots, n_q^k$ can be found to satisfy the inequality:

$$-3 < \mu - n_i^k + \frac{k\sigma}{2^m} < \frac{3\sigma}{2^m} \quad (1 \le i \le q). \tag{16}$$

The corresponding values from mother wavelet mask can then be taken to calculate:

$$\alpha_i^k = \frac{2^{m/2}}{\sigma} F(\frac{\rho_i^k}{2^m}),$$

where $\rho_i^k = \left[(\mu - n_i^k)2^m + k\sigma\right] \quad (1 \le i \le q)$

and

$$\frac{c_{-m,k}}{\sqrt{\alpha}} - \sum_{i=1}^{q} c_{ni}\alpha_i^k. \tag{17}$$

Decomposing the re-normalized signal $\frac{c_{-m,k}}{\sqrt{\alpha}} \quad (k \in Z)$ according to the conventional DWT, will result in the entire DMAW filter basis for different scales:

$$\frac{c_{-m+1,k}}{\sqrt{\alpha}}, \frac{d_{-m+1,k}}{\sqrt{\alpha}}, \frac{c_{-m+2,k}}{\sqrt{\alpha}}, \frac{d_{-m+2,k}}{\sqrt{\alpha}}, \ldots., \frac{c_{0,k}}{\sqrt{\alpha}}, \frac{d_{0,k}}{\sqrt{\alpha}} \tag{18}$$

4. Experimental results

Figures 1, 2, and 3 show the experimental results for the application of the proposed filter banks to a noisy sinusoidal signal. As is evident from these figures, a signal can be decomposed in as many levels as desired by the application and allowed by the computational constraints. Levels shown from top to bottom represent the coarse to detailed components of the original signal.

Once the signal is decomposed to its components, it is easy to do away with pieces that are not needed. For instance, noise, which is the lower most signal in Figure 1 can be totally discarded. On the other hand, if compression is necessary, all but the coarse component (upper most element, below the original signal) can be kept and the rest of the modules discarded. This signal alone is a fairly good approximation of the original signal. Figure 2 shows the thresholds and coefficients of the signal being filtered.

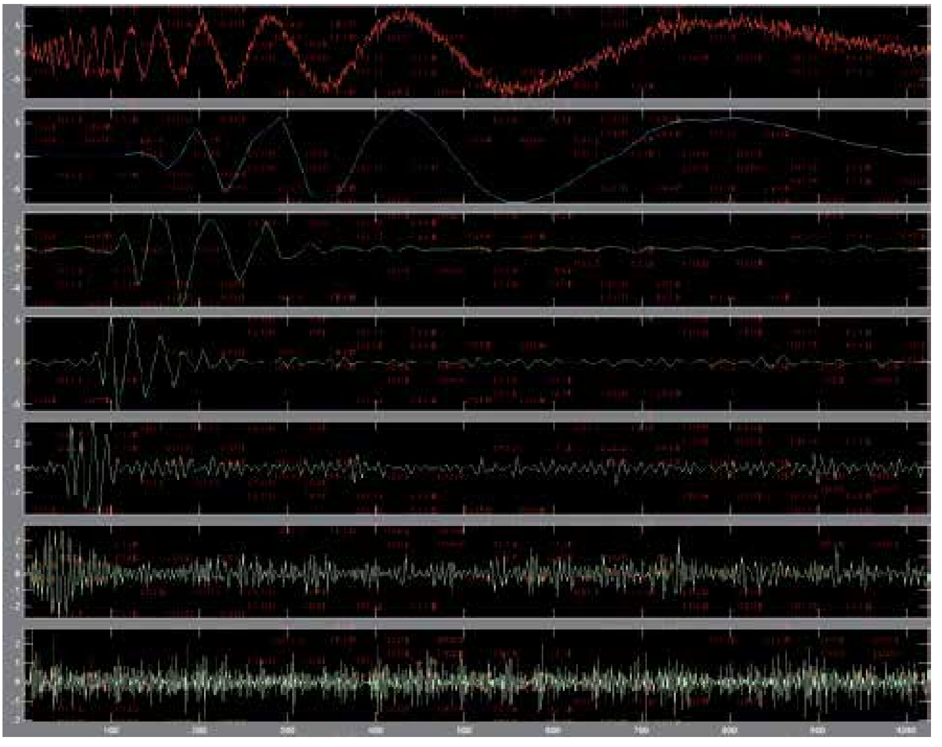


Figure 1. Decomposed signal showing all the components of a mixed sine wave with noise

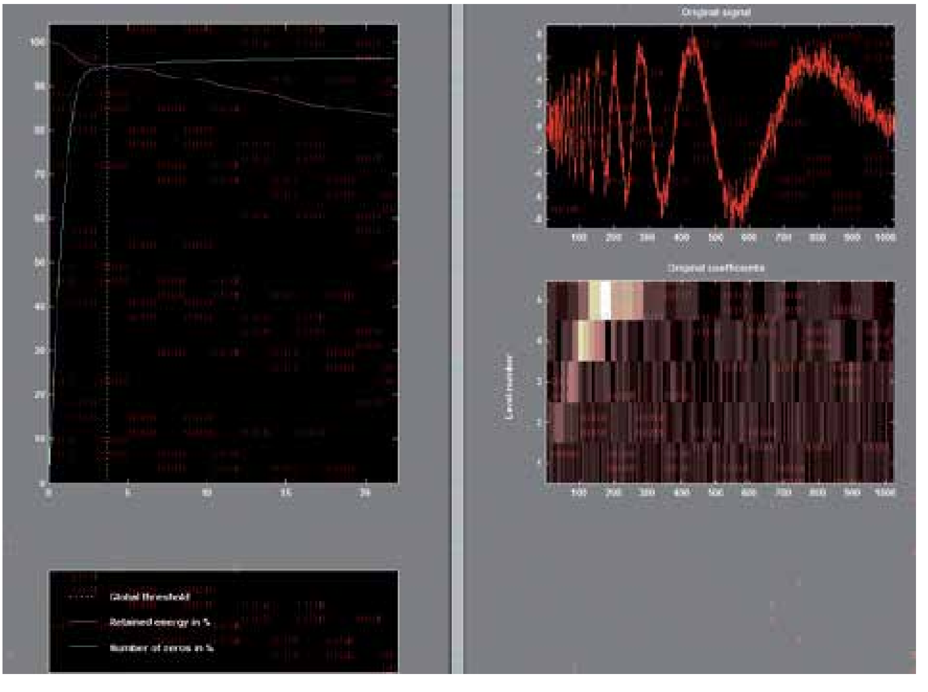


Figure 2. Threshold and coefficients of the decomposed signal

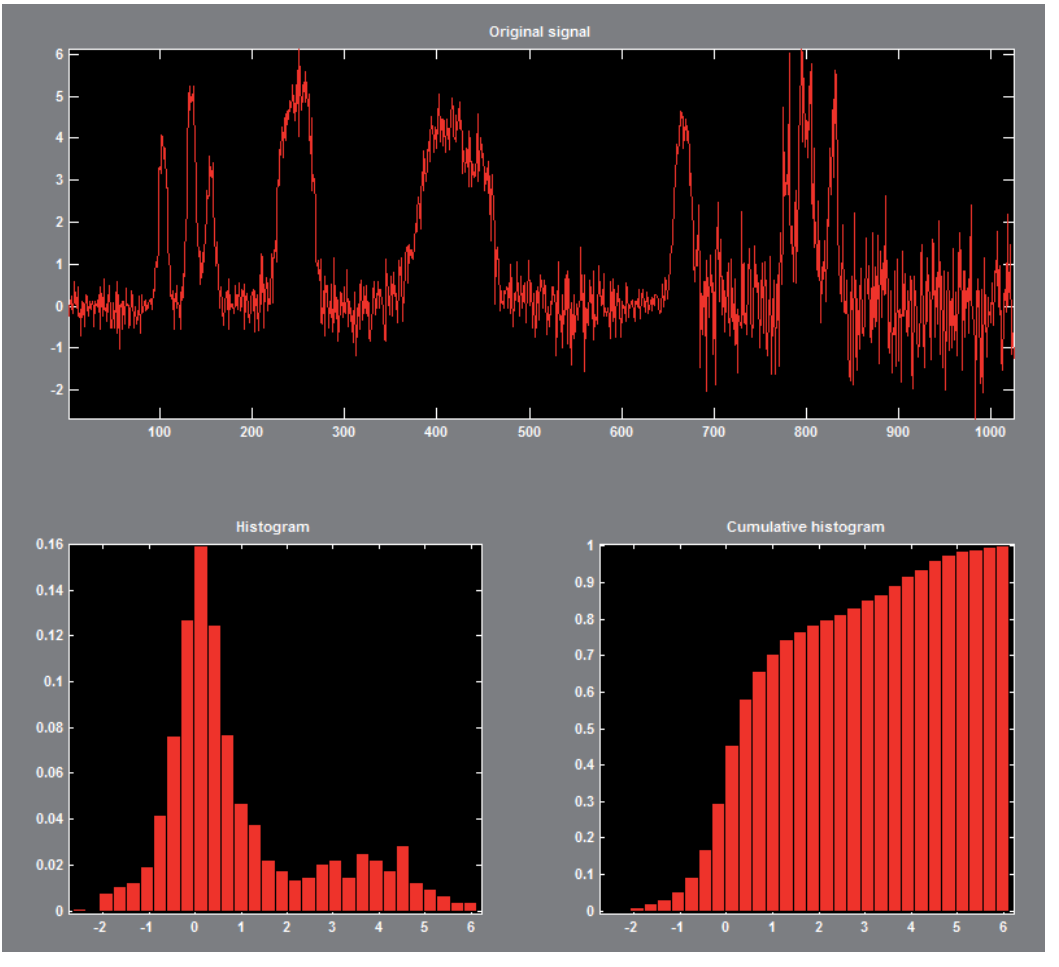


Figure 3. Histogram and cumulative histogram of the signal

Figure 3 shows the histogram (frequency of components distribution) of the signal. For comparison purposes the same filter banks have also been applied to a quad-chirp signal with noise and the results are shown in Figures 4-9. The denoised and compressed versions of the signal have been computed and plotted. In each case the coefficients that have remained intact for denoising and compression have also been displayed. Finally, in Figures 7-25 the histogram for the denoised and compressed quad-chirp, auto-regressive, white noise, and step signals have been compared to the original signal. The effectiveness of the proposed filter banks and their capability to maintain the important components of the original signal is evident in these figures.

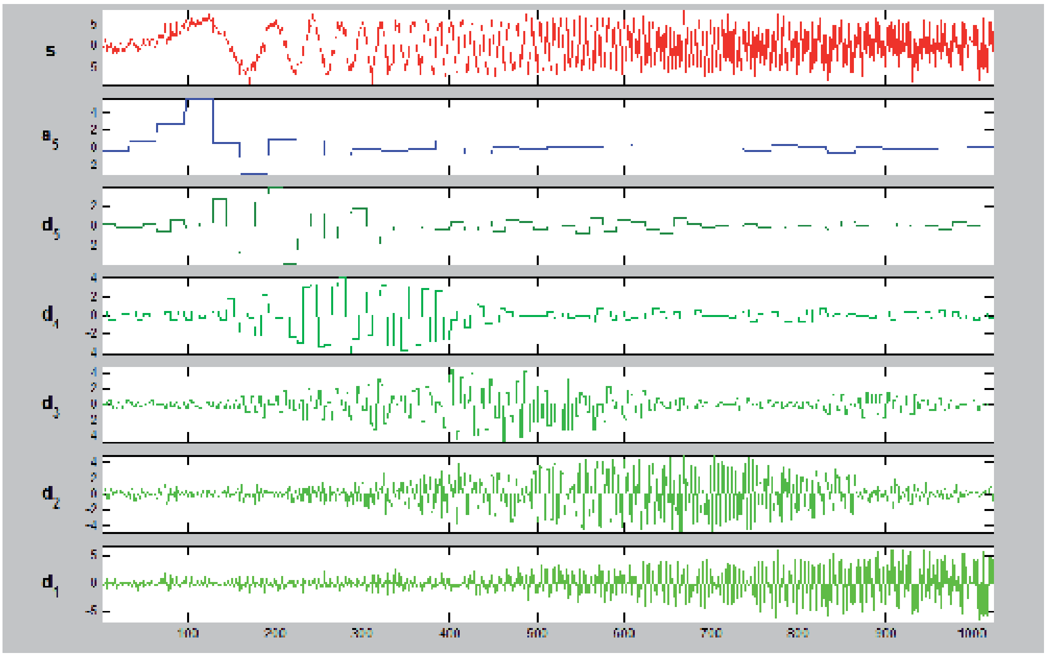


Figure 4. Decomposed signal showing all the components of a quad-chirp wave with noise

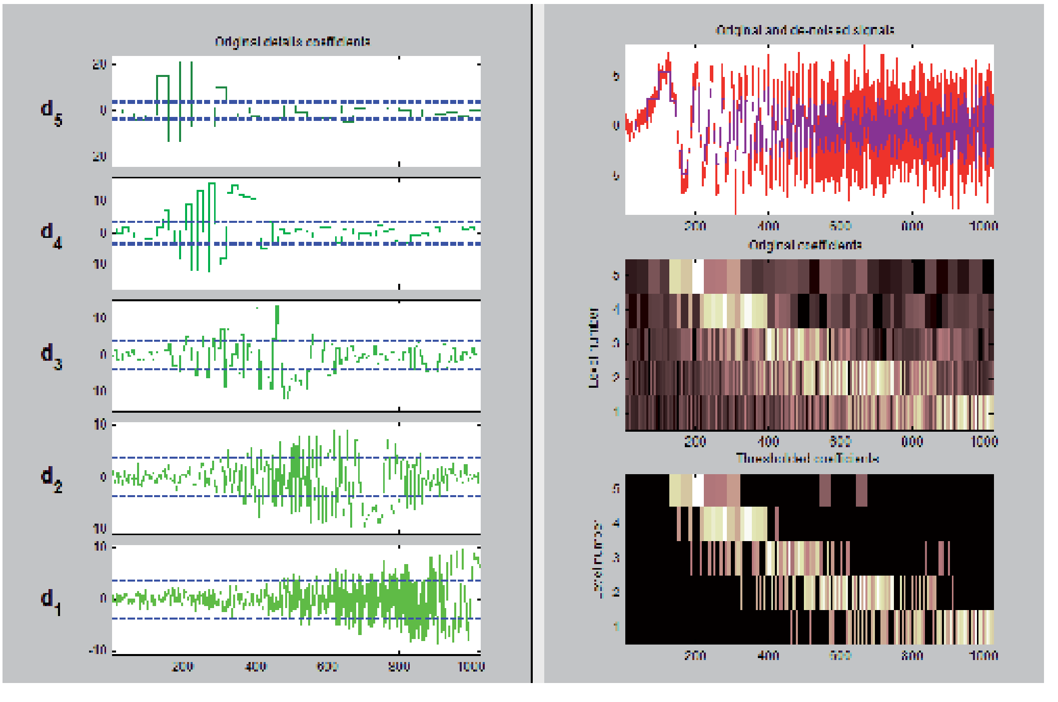


Figure 5. Original and denoised signal with original and thresholded coefficients

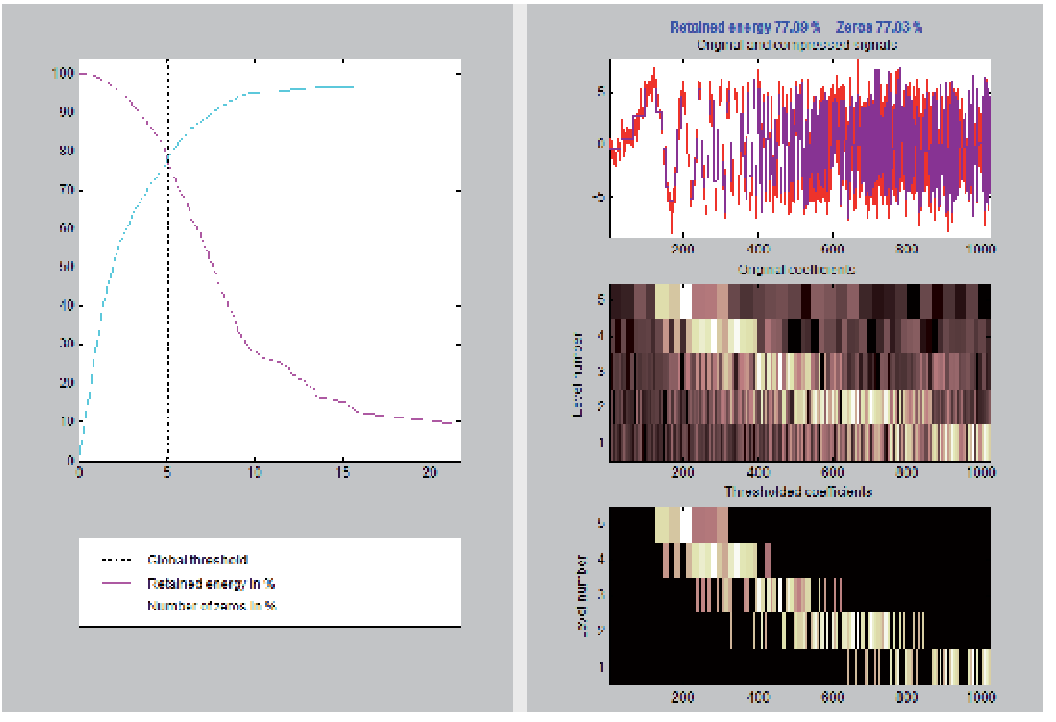


Figure 6. Threshold and coefficients of the decomposed signal showing retained energy and number of zeros

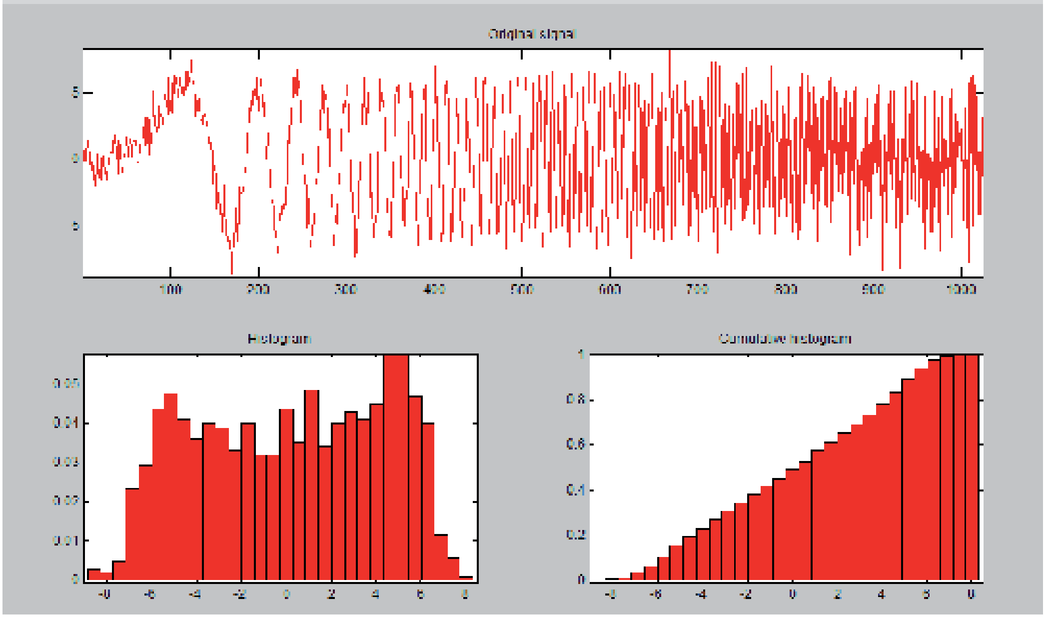


Figure 7. Histogram and cumulative histogram of the original quad-chirp signal

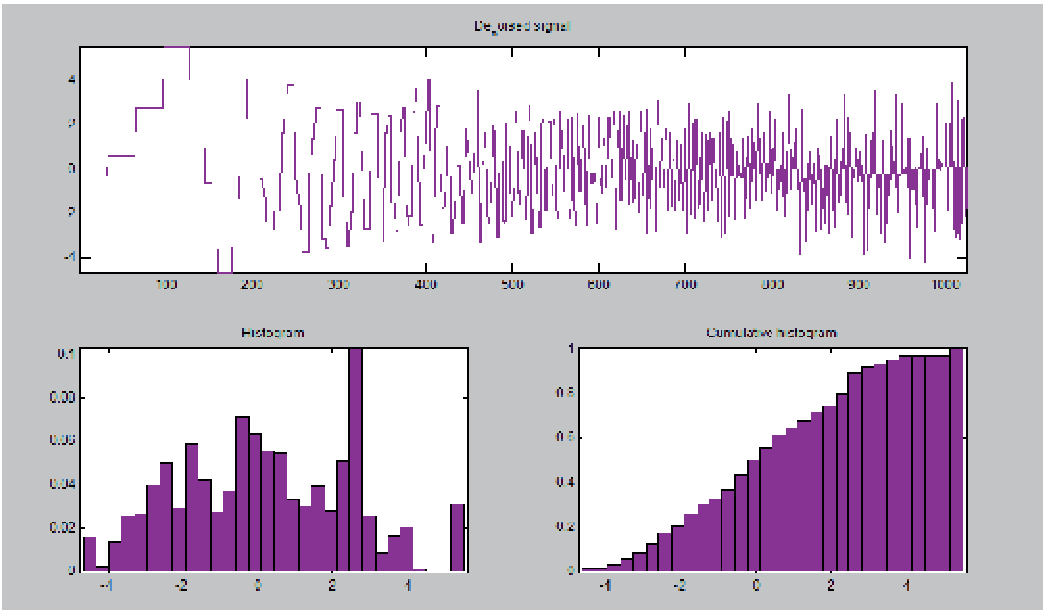


Figure 8. Histogram and cumulative histogram of the denoised quad-chirp signal

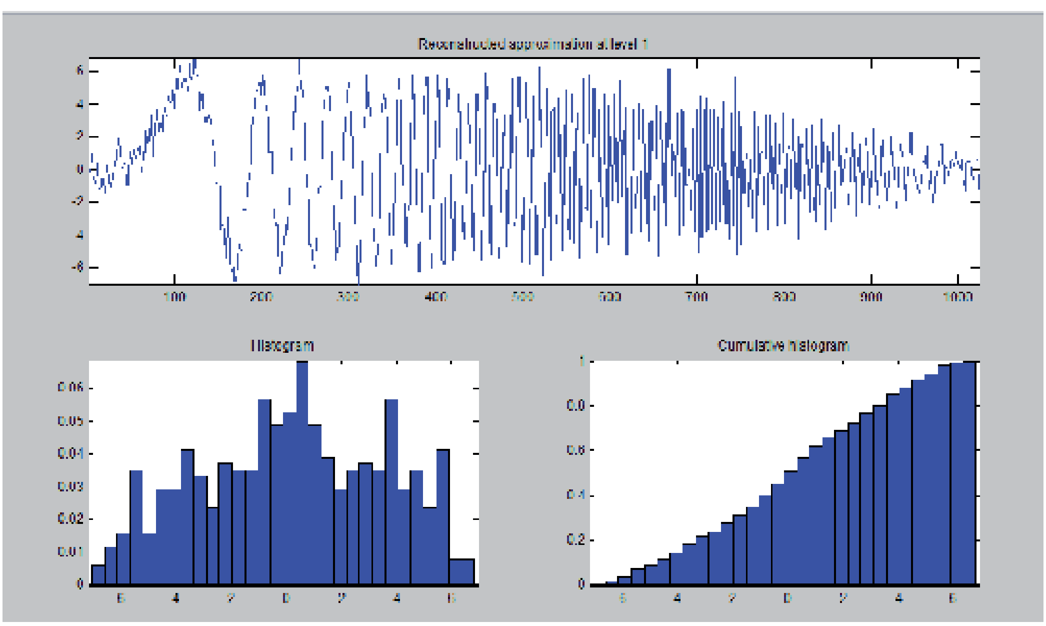


Figure 9. Histogram and cumulative histogram of the compressed quad-chirp signal

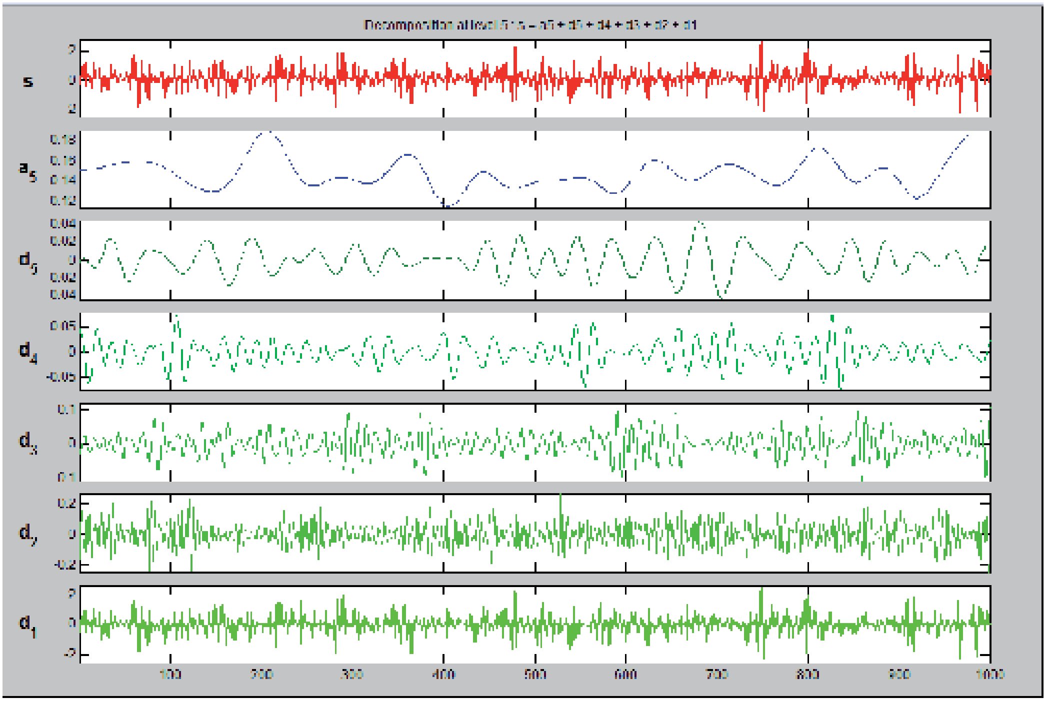


Figure 10. Decomposed signal showing all the components of an auto-regressive wave with noise

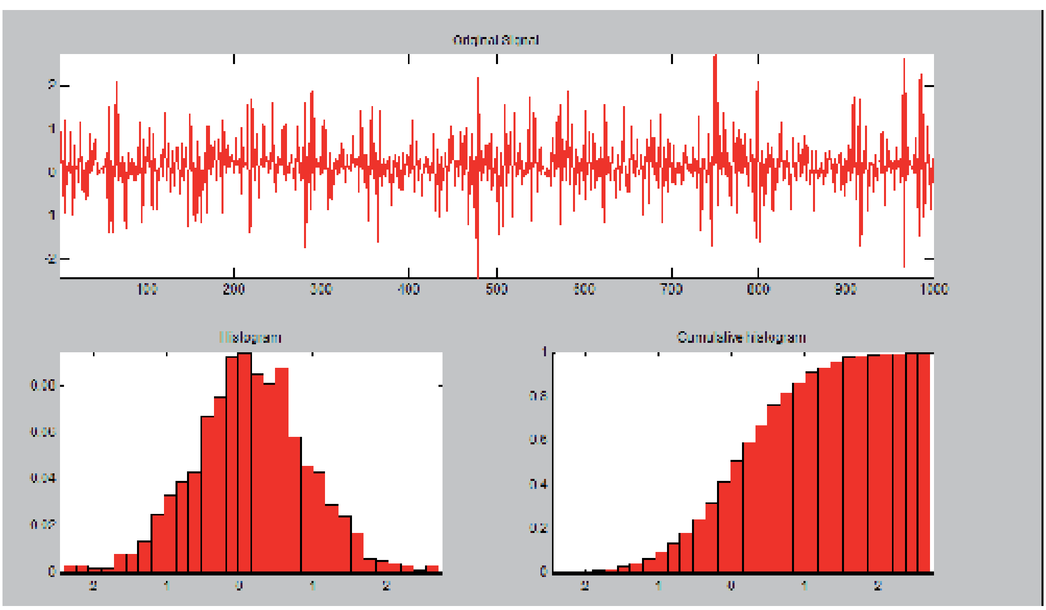


Figure 11. Histogram and cumulative histogram of the original auto-regressive signal

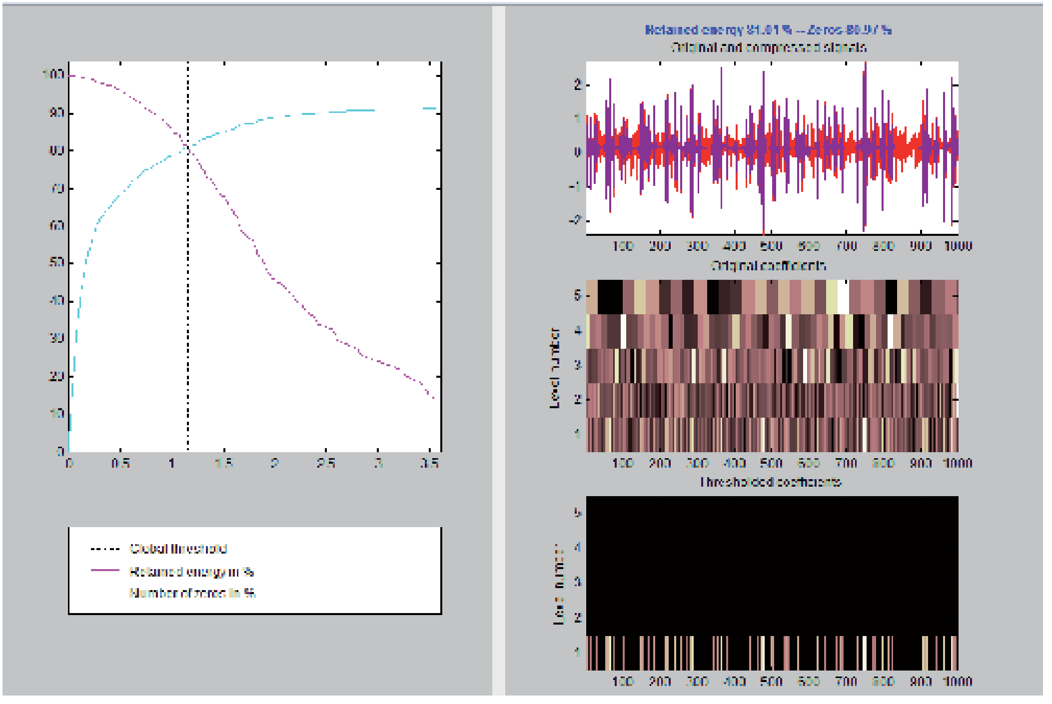


Figure 12. Threshold and coefficients of the decomposed signal showing retained energy and number of zeros

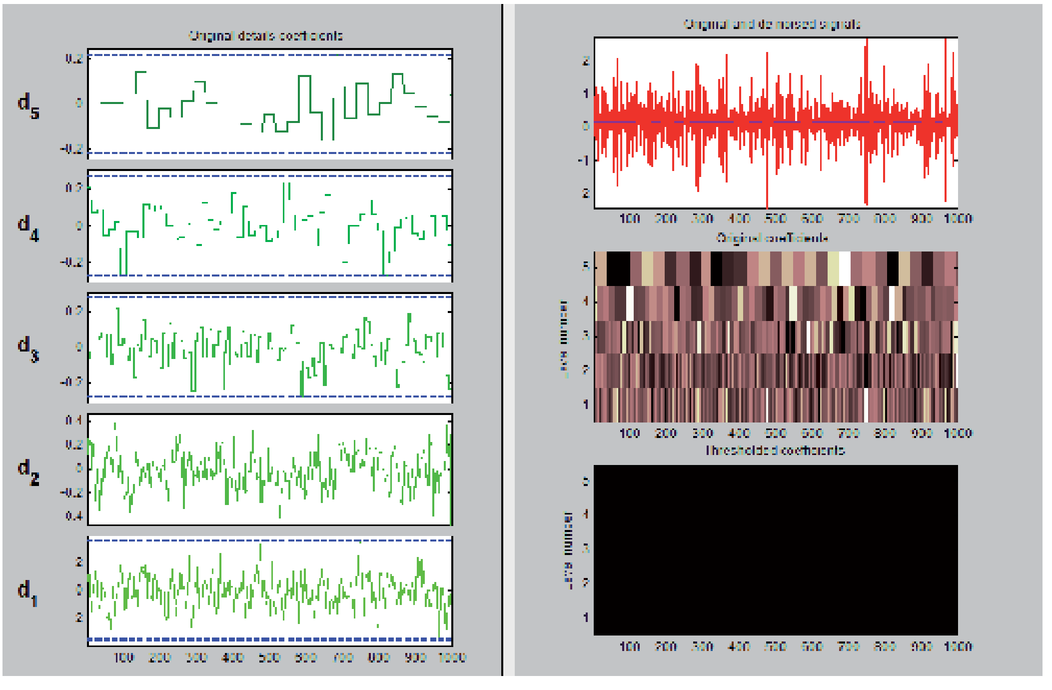


Figure 13. Original and denoised signal with original and thresholded coefficients

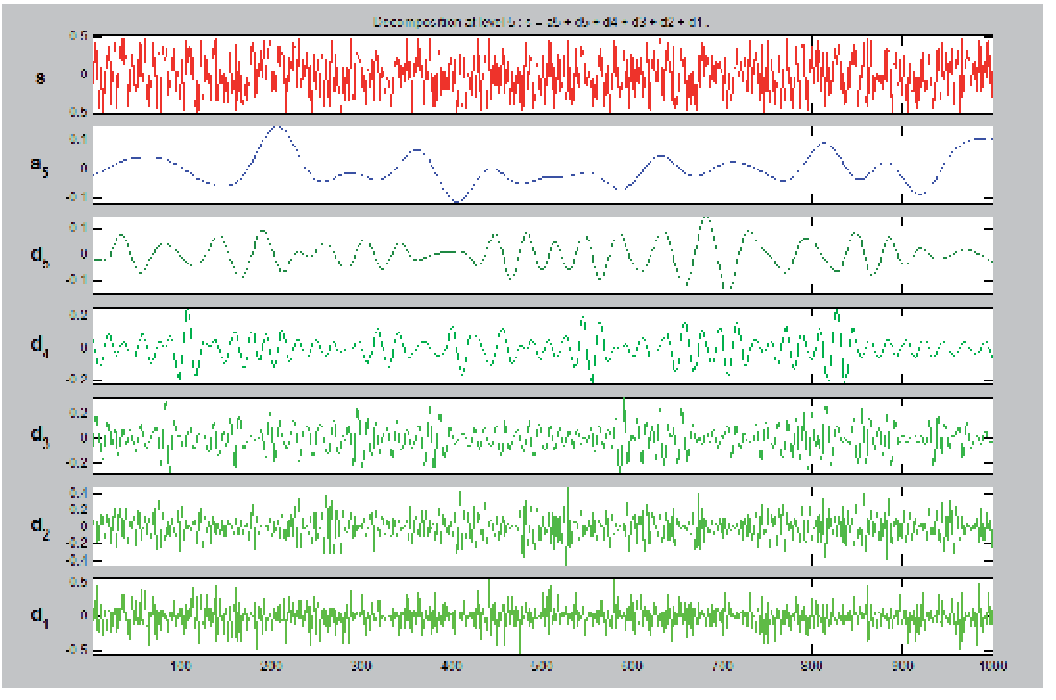


Figure 14. Decomposed signal showing all the components of white noise

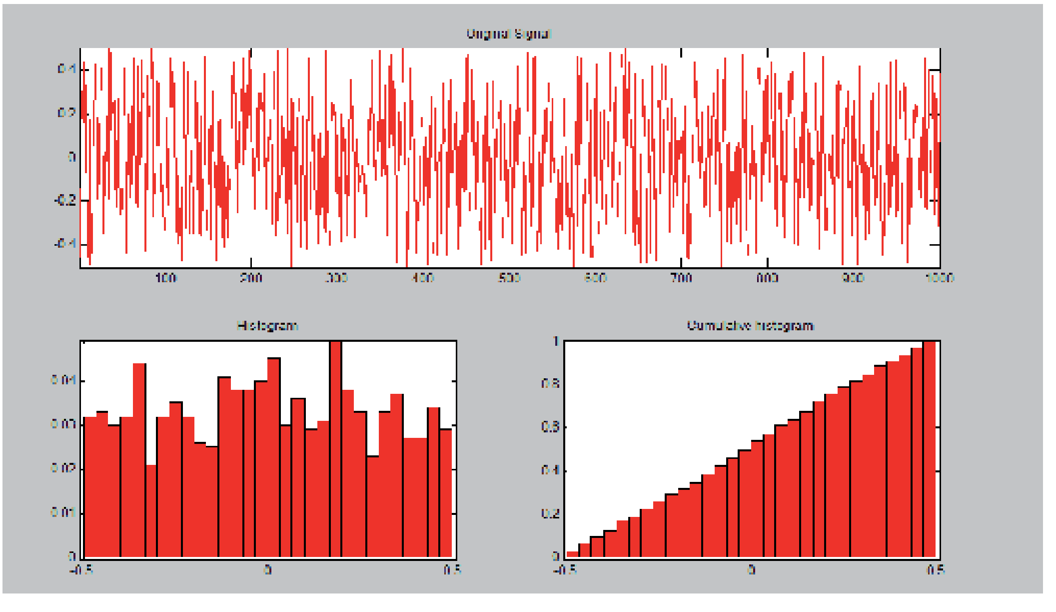


Figure 15. Histogram and cumulative histogram of the original white noise signal

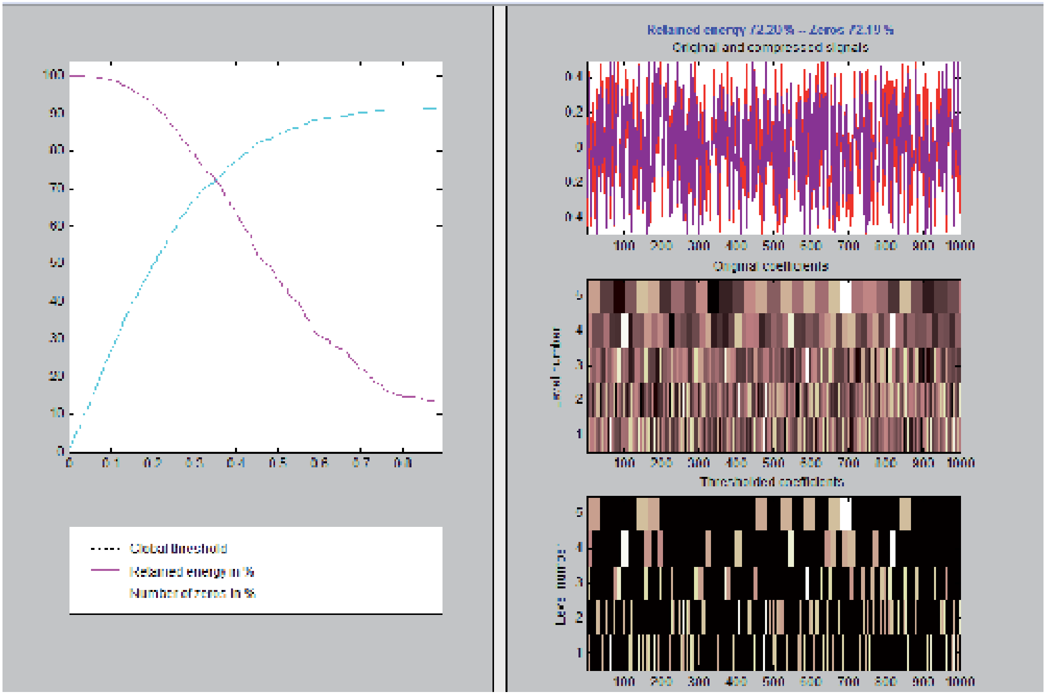


Figure 16. Threshold and coefficients of the decomposed signal showing retained energy and number of zeros

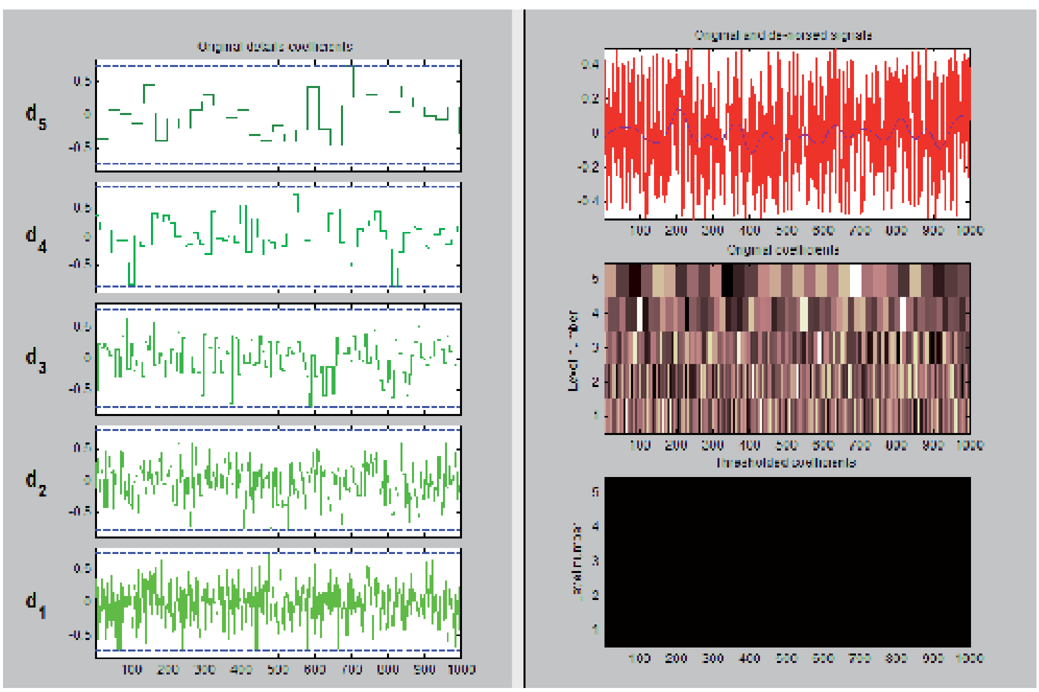


Figure 17. Original and denoised signal with original and thresholded coefficients

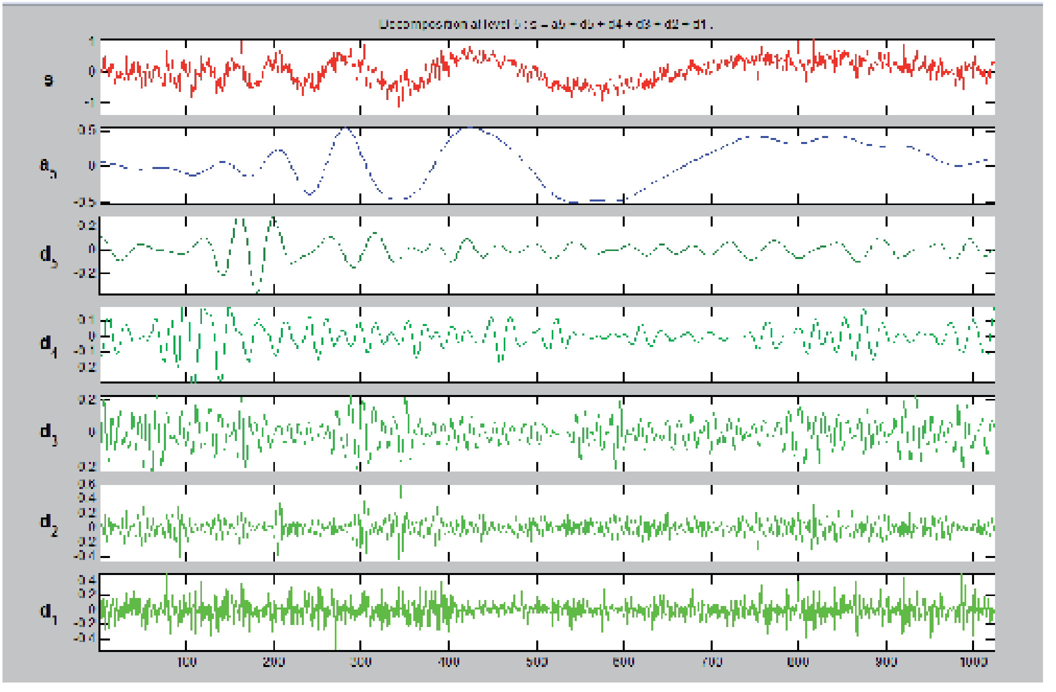


Figure 18. Decomposed signal showing all the components of a doppler wave with noise

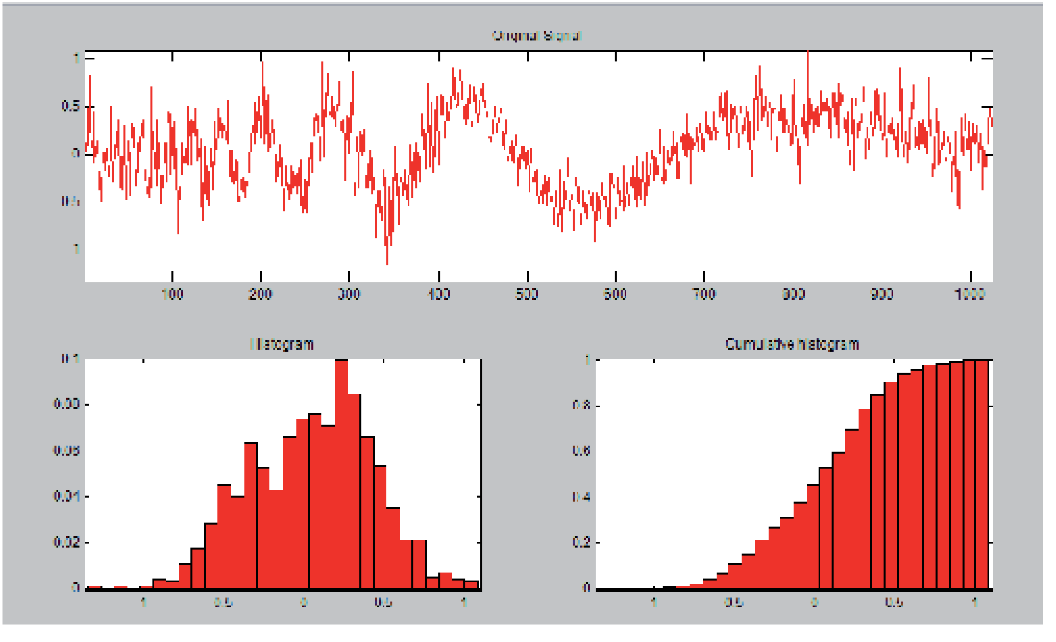


Figure 19. Histogram and cumulative histogram of the original doppler signal

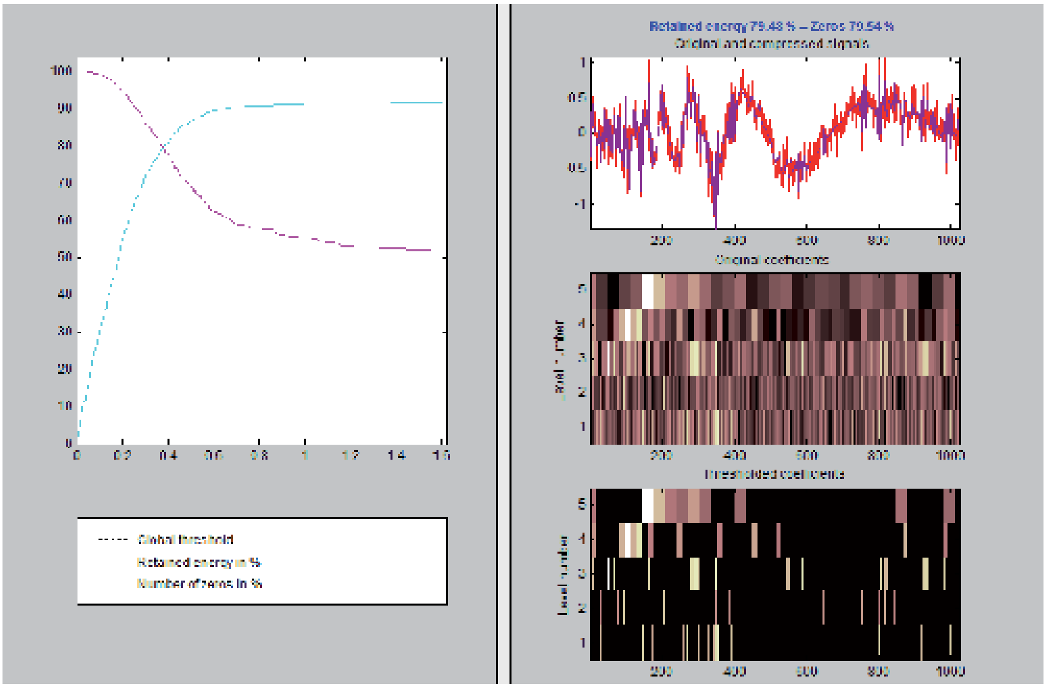


Figure 20. Threshold and coefficients of the decomposed signal showing retained energy and number of zeros

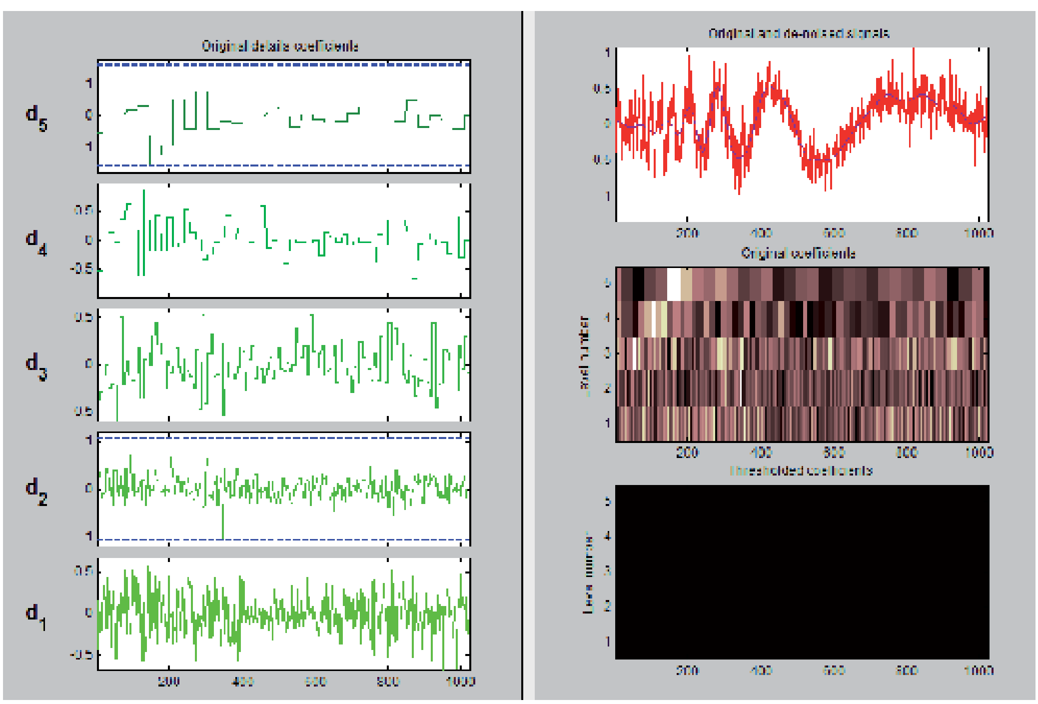


Figure 21. Original and denoised signal with original and thresholded coefficients

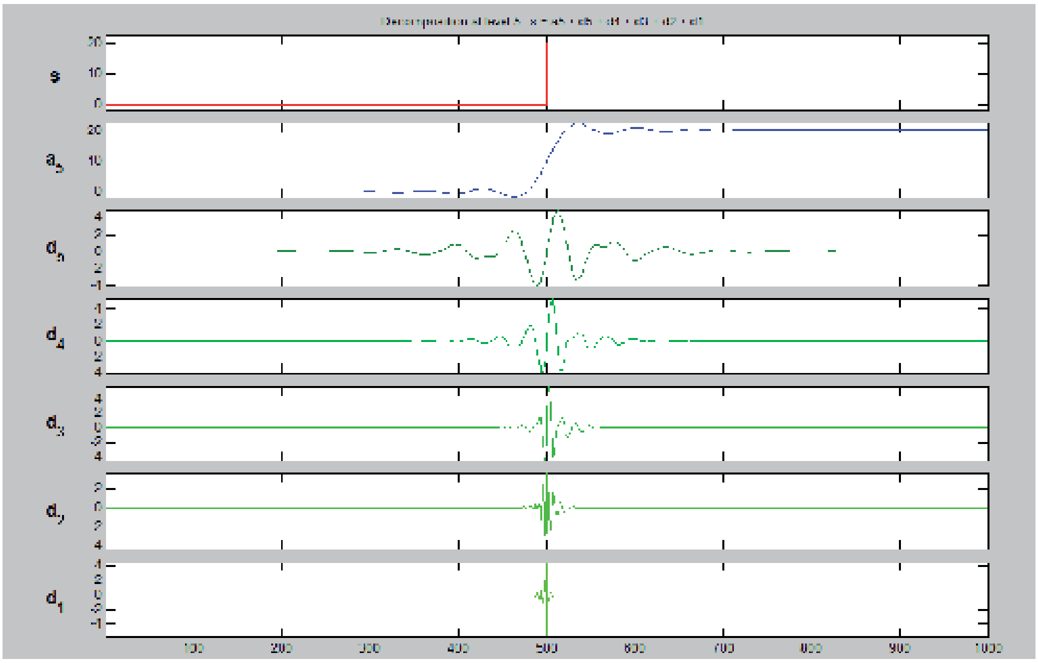


Figure 22. Decomposed signal showing all the components of a step signal

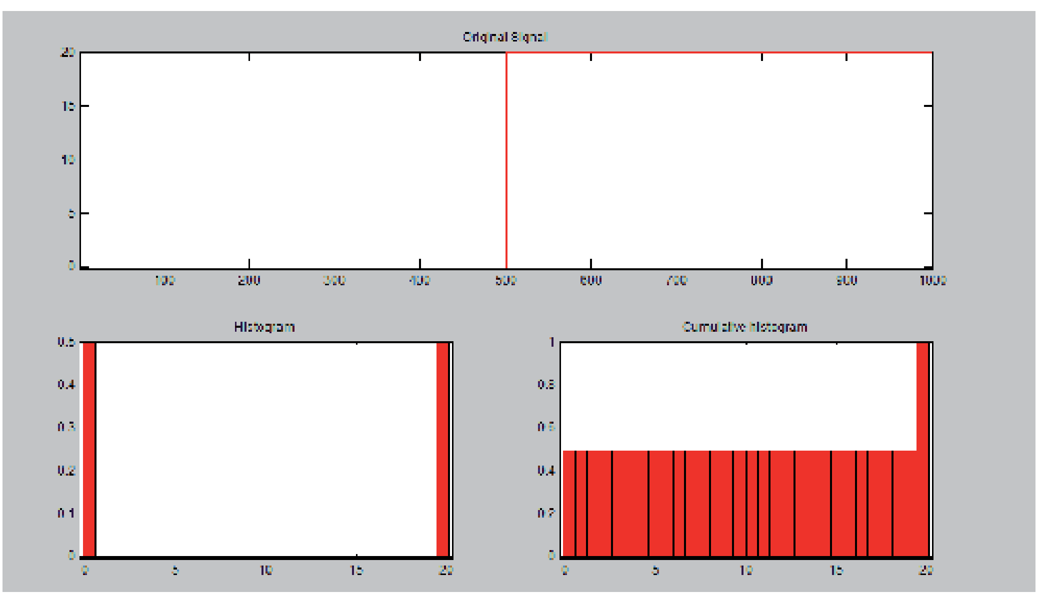


Figure 23. Histogram and cumulative histogram of the original step signal

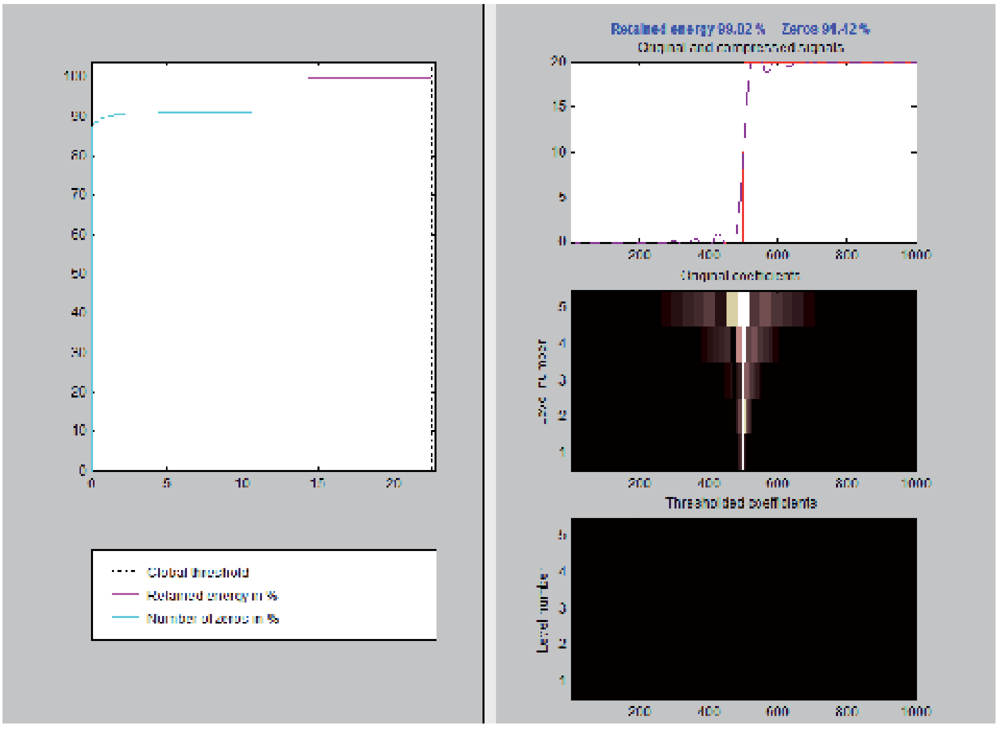


Figure 24. Threshold and coefficients of the decomposed signal showing retained energy and number of zeros

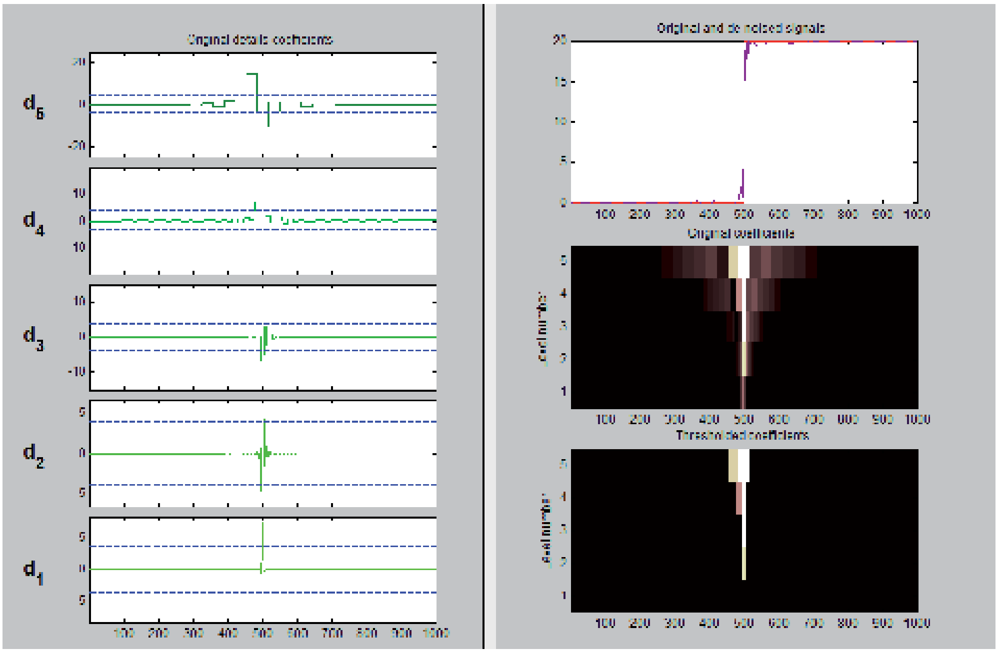


Figure 25. Original and denoised signal with original and thresholded coefficients

5. Conclusions and future work

As expected from the theory, the DMAW filters performed well under noisy conditions in a wireless sensor network. The decomposed signal could be easily freed up from noise and reduced down to its coarse component only. This could be reduction by several orders of magnitude in some cases. Future plans include the application of these filters to fused datasets and comparison between the two approaches. Additionally, the results of these study can be used in the decision making stage to realize the difference this approach can make in speed and efficiency of this process.

Future work will address issues such as characterizing the parameters for simulation and modeling of the proposed filter for WSN; showing how complex examples with correlated sensor data will be filtered for redundancy; comparing the proposed approach with other similar approaches and giving comparative results to support the claimed advantages, both theoretically and experimentally.

Author details

Ehsan Sheybani
Virginia State University, Petersburg, VA, USA

6. References

Abdallah, A., Wolf, W., "Analysis of Distributed Noise Canceling", *Distributed Frameworks for Multimedia Applications, 2006. The 2nd International Conference on*, May 2006, pp. 1–7.

Closas, P., Calvo, E., Fernandez-Rubio, J.A., Pages-Zamora, A., "Coupling noise effect in self-synchronizing wireless sensor networks", *Signal Processing Advances in Wireless Communications, 2007. SPAWC 2007. IEEE 8th Workshop on*, 17-20 June 2007, pp.1-5.

Cohen, I., S. Raz, and D. Malah, "Shift invariant wavelet packet bases," in *Proc. 20th IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, May 8–12, 1995, pp. 1081–1084.

Coifman, R.R. and M. V. Wickhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.

Daubechies, I., "Ten lecture on wavelet," in *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, PA: SIAM, 1992.

Liang, J. and T.W. Parks, "A translation invariant wavelet representation algorithm with applications," *IEEE Trans. Signal Processing*, vol. 44, pp. 225–232, Feb. 1996.

Mallat, S., "Zero-crossings of a wavelet transform," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1019–1033, July 1991.

Mallat, S. and S. Zhang, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, July 1992.

Pescosolido, L., Barbarossa, S., Scutari, G., "Average consensus algorithms robust against channel noise", *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, 6-9 July 2008, pp. 261–265.

Schizas, I.D., Giannakis, G.B., ”Zhi-Quan Luo Optimal Dimensionality Reduction for Multi-Sensor Fusion in the Presence of Fading and Noise”, *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 14-19 May 2006, Vol. 4, pp. IV – IV.

Son, S.-H., Kulkarni, S.R., Schwartz, S.C., Roan, M., “Communication-estimation tradeoffs in wireless sensor networks”, *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP apos;05). IEEE International Conference on*, Volume 5, Issue , 18-23 March 2005, pp. 1065-1068, Vol. 5.

Yamamoto, H., Ohtsuki, T., “Wireless sensor networks with local fusion:, *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, 28 Nov.-2 Dec. 2005, Vol. 1, pp 5.

Performance Analysis of a Compression Scheme for Highly Dense Cluster-Based Wireless Sensor Network

Sofiane Moad, Mario E. Rivero-Angeles, Nizar Bouabdallah,
Bruno Sericola and Yassine Hadjadj Aoul

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48370

1. Introduction

Wireless Sensor Networks (WSNs) consist of spatially distributed autonomous devices, called sensors, that communicate in a wireless manner. Sensors cooperate together to monitor physical or environmental conditions such as temperature, sound, vibration or pressure [1]. This technology has been originally developed for military applications such as battlefield surveillance [2]. Recently, WSNs have been deployed in many other civilian application areas. It includes environment applications such as fire detection in forest monitoring and health-care applications like monitoring the patient status.

Sensor nodes are expected to be of tiny size. Therefore, the size of every sensor's components, such as the power source, processing and data storing memory, are expected to be also very limited. In addition of the physical characteristics, a large number of sensors are often deployed in hostile environments, where the human intervention is difficult if not impossible, for example inside a volcano. Hence, in these networks it is not practical to perform maintenance operations, such as changing batteries on deployed sensor nodes. This requires sensors to be able to self-organize, self-configure and should optimize the energy consumption to maximize the network's lifetime. The network lifetime in WSNs refers to the period of time from the deployment of the sensor nodes to the instant when the network is considered unusable [3].

In terms of energy consumption, sensors consume energy for three main reasons: data sensing, data processing and wireless data communicating. Wireless communication refers to data transmission and reception. Among these three operations, the most power-consuming task is data transmission. Approximatively 80% of power consumed in each sensor node is used to transmit data [4]. One field of research, aimed at extending the lifetime for WSNs, is data compression. In general, by applying a suitable compression scheme, power consumption can be reduced during the transmission and processing stages thus extending network lifetime. Also, by reducing the data packet size, less bandwidth is required to send and receive data [4]. In addition, to further enhance energy efficiency, cluster-based communication schemes are widely used in WSNs [5]. In view of this, we propose to take advantage of both data compression and clustering in order to further reduce energy consumption in the network. Indeed, the compression and clustering strategies are not supposed to work independently but rather the network design has to consider both simultaneously. Therefore, by enabling both compression and clustering schemes in WSNs, energy consumption would be greatly enhanced.

In this chapter, we propose a complete analysis of a new Compression Cluster-based scheme in a Spatially-Correlated Region (CC_SCR) for event-driven applications in WSNs. As WSNs are typically densely-deployed over a sensor field [6], sensor nodes are typically very close to each other. Contrary to continuous monitoring applications, in Event Detection-Driven (EDD) applications, active nodes are concentrated in a relatively small area. Therefore, readings from these nodes are expected to be quite similar. Building on this, we propose a clustering scheme that exploits the spatial correlation of sensed data among the nodes to reduce the size of the data packets that will be sent. Specifically, in the proposed scheme, Cluster Members (CMs) send only the differences between their readings and a reference data which corresponds to the value sensed by the selected Cluster Head (CH). As such, one of the proposal's main issue is to select as CH the node that reduce the average packet size in the cluster. Note that, in this chapter, we complement our previous work published in [7] with several simulation results and analytical analysis.

The main contributions to this chapter are:

1. The CC_SCR compression cluster-based protocol for WSNs is proposed. It exploits the spatial correlation of the sensed data in order to reduce energy consumption.
2. An analytical energy consumption model for comparison to both classical and single hop schemes is developed.
3. The CC_SCR is implemented in TinyOS [8] for simulation analysis, in order to prove the potential benefits of CC_SCR in future applications of real WSNs deployments.

The remainder of this chapter is organized as follows. Section 2 presents a background of research related to this work, while Section 3 exposes the network model. Section 4 shows the analytical results regarding the energy consumption and network lifetime, while Section 5 shows the simulation results. Finally, the chapter concludes in Section 6 with a summary of the main advantages of the proposed scheme.

2. Related work

We review some of the related works regarding the compression and the clustering schemes in Subsections 2.1 and 2.2, respectively.

2.1. Compression schemes

In the literature, there has been an increased interest in studying compression algorithms for WSNs. On the other hand, many of these compression algorithms have been proposed for classical networks, which are not suitable to be deployed in the WSNs context [9, 10]. The main reason is the limited memory size of sensor nodes, for example, the size of bzip2 is $219KB$ and the size of LZO is $220KB$. Another reason is the limited processor speed of sensor nodes which is around $4 - 8MHz$. Thus, embedding classical data compression schemes in these tiny nodes is very difficult and it is necessary to design a low-complexity and small-size data-compression algorithm for sensors. We review in the following some of these compression techniques. A more detailed description of compression methods can be found in [11].

One compression algorithm for WSNs is the coding-by-ordering data-compression technique [12]. In this algorithm, when data is combined at an aggregation node, some of these data are implicitly transmitted. The main idea behind this technique is to replace the data transmission of certain nodes by the order in which the aggregated packets are placed by the aggregator. For example, consider five nodes (n_1, n_2, n_3, n_4, n_5) that send data to their aggregator node, n_a and suppose that data value of each node can be any integer from a range of 0 to 23. When the aggregator node receives a value from each node n_1, n_2, n_3, n_4, n_5, the order of transmission of first four nodes n_1, n_2, n_3, n_4 determines the value of n_5 implicitly. Indeed, there are $4! = 24$ possible ways of ordering these data packets.

The pipelined in-network compression algorithm is discussed in [13]. The main idea behind this technique is to combine data in order to make them smaller than the original size. After the aggregator collects data from differents nodes, it is stored for a certain amount of time. Data packets are then combined into one packet to minimize data transmission. For example, consider that each data packet has the following form: $<$ *data value*, *node ID*, *timestamp* $>$. Then, the compressed data packet has the following form: $<$ *shared prefix*, *suffix list*, *node ID list*, *timestamp list*$>$. The *shared prefix* field, i.e. the most significant bits, is the same for all the measured values. The suffix list field expresses the list of measured values excluding the *shared prefix* part. The *node ID list* is the list of node identifiers and the *timestamp list* is the list of timestamps. One advantage of this simple compression scheme is that the shared prefix system can be used for both *node ID* and *timestamp* fields. By doing so, more data compression can be achieved.

The distributed compression scheme proposed in [14] uses a side information to encode a source information. For example, if there are two data sources: X and Y, which are correlated and chosen from a discrete alphabet, then X can be compressed at the theoretical rate of its conditional entropy $H(X|Y)$. The receiving node maintains the cosets and can decode X knowing Y's codevector, and with partial information from the source X.

The main advantage of the *CC_SCR* algorithm, compared to those previously mentioned is that *CC_SCR* takes into account the physical characteristics of the sensed data in order to compress data and also considers clustering communication in order to reduce energy consumption.

2.2. Clustering schemes

In addition to the compression techniques, there has been an increased interest in studying energy efficient clustering algorithms and extensive clustering algorithms have been proposed for WSNs. Hereafter, we briefly review the most relevant energy efficient clustering algorithms. For more details, the reader can review [15], [16], [5], and [17].

Hybrid Energy-Efficient Distributed clustering (HEED) [18] protocol operates in two main phases: the set-up phase where clusters are formed and the steady phase where the sensor nodes transmit their data using the Time Division Medium Access (TDMA) frames. HEED set-up phase operates in three sub-phases. The first sub-phase is the initializing. Nodes exchange *hello* messages to discover their neighborhoods. The second sub-phase consists of a competition process. The third sub-phase is the finalizing and it allows nodes to join their corresponding CH based on the connectivity degree.

Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [19] steady phase consists of a formation of chains instead of clusters. In the chain formation, the Base Station (BS) and sensor nodes are connected via a chain using a greedy algorithm. One of the nodes, in the chain, is selected by turns to represent the head. In data gathering phase, each node delivers the sensing data to the nearest neighbor node until the data reach the head node which aggregates and delivers the sensing data to the BS.

In [20], the authors proposed an Energy Efficient Clustering Scheme (EECS) protocol. In this protocol, CH candidates compete for the ability to elevate to a CH with a certain probability. This competition involves candidates broadcasting their residual energy to neighboring candidates. If a given node does not find a node with more residual energy, it becomes a CH.

The main difference between the aforementioned clustering algorithms and the CC_SCR algorithm is that nodes use the physical characteristic of the sensed data to elects CHs. The benefits of our proposed scheme is that nodes use the compression scheme to reduce the energy consumption.

In this chapter, we combine the benefits of using both the clustering and the compression techniques to reduce the energy consumption in the network. Indeed our proposal scheme takes into account the characteristics of the physical surveilled event and also takes advantage of the energy unconstrained of the Base Station (BS) which participates into the CH selection. Specifically, our proposal scheme takes advantage of the fact that the nodes that sense a certain event are usually very close to each other (which entails a high correlation between sensed data) in order to reduce the size of data packets communicated through the network. The BS then selects an efficient CH that minimize the data transmission over the network. To the best of knowledge this is the first clustering protocol, that takes into consideration the physical characteristics of the environment to elect energy efficient clusters and therefore implement the compression scheme to reduce the energy consumption.

3. Network model

We consider an event-driven WSN consisting of N sensors deployed over a vast field as in Figure 1. We denote the *i-th* sensor node as n_i and the corresponding sensor node set as $\{n_1,$

$n_2,...,n_N\}$. Some assumptions concerning the sensor nodes and the underlying network model are now presented:

- Nodes are uniformly distributed in an $A \times A$ area with (x, y) coordinates. Nodes are homogenous and all have the same capabilities. A unique identifier ID is assigned to each node.
- Sensor nodes and the sink node are all stationary after deployment.
- Nodes have two power controls to vary transmission power, which depends on the distance to the receiver [21]. Each node, n_i, can reach any other node with a transmission range, R_c. The sink can be reached with a transmission range, $R_t > R_c$.
- CHs use the average operation as the aggregation technique in order to eliminate data redundancy.

We consider event detection driven wireless sensor applications. The center of the event is located in a random uniformly-distributed point with coordinates (x_{event}, y_{event}) within the network's area. The range of the event, i.e., the area range where sensors can detect the event, is R_event meters, where $R\_event \in [1, A]$. We also suppose that an event has a duration of T_event seconds. In addition, only sensor nodes within R_event range are considered as active nodes in the network and they are the only nodes performing as the source of the detected event. The rest of the nodes in the system are not considered in our analysis as they do not participate in data reporting. Additionally, in our model, only one event can be active inside the system area and the data value C at the center of the event is constant, i.e., the stationary model in which the measured data do not change during the T_event seconds that the event is active.

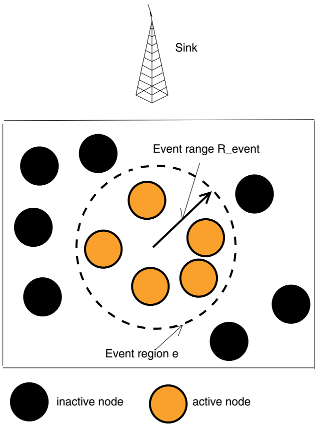


Figure 1. Event-driven application in WSNs.

A cluster-based WSN is considered where only one CH is elected for each event. The clustering process is triggered whenever an event is sensed by the nodes inside the event area.

The spatial correlation of the data from the different active nodes has been modeled on previous works in the area according to the following models:

1. Diffusion propriety [22].
2. Data is jointly Gaussian, with the correlation being a function of the distance [23].
3. Data is a function for their joint entropy [24].
4. Correlation is calculated from realistic environmental monitoring and testbeds [25].

In this chapter, we use diffusion property to model spatially-correlated data [26]. The model considered in this chapter is the same as in [22] in which the data reading at a distance d from the center of the event is $D = C/(d+1)^{\alpha}$, where C is a constant representing the value at the center of the event, and α is the diffusion parameter, which depends on the particular environment and phenomenon surveyed, e.g., for light $\alpha = 2$, heat $= \alpha \simeq 1$.

Figure 2 shows the data reading using the aforementioned model, with different values for α, and $C = 250$. On one hand, when $\alpha \geq 1$, we observe a relatively big difference between the value sensed at the center of the event and the values observed at distance d far away from the center of the event. On the other hand, when $\alpha < 1$ ($\alpha = 0.1, 0.01$ and 0.001), the data readings away from the center of the event are very similar. In our study, we are interested, specifically, in the types of event where data values are highly correlated.

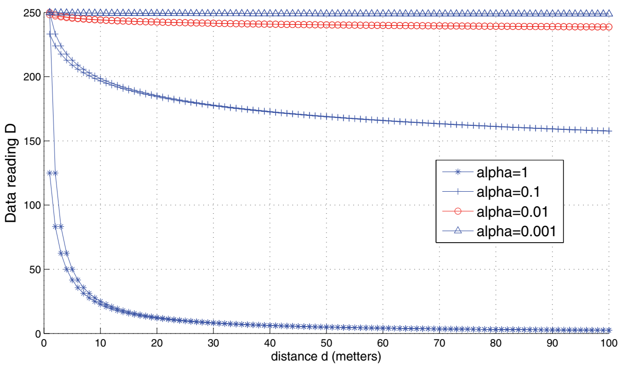


Figure 2. Variation in data reading with distance from the center of the event.

We use Henizelman's energy consumption model [27]. Specifically, the energy consumed to transmit a message at a distance d is given by

$$E_{tx}(sz,d) = \begin{cases} sz \times E_{elec} + sz \times E_{fs} \times d^2, \text{ if } d \leq d_0. \\ sz \times E_{elec} + sz \times E_{mp} \times d^4, \text{ otherwise.} \end{cases} \tag{1}$$

where sz is the data packet size in bits, E_{elec} is the energy consumed due to the transmitter/receiver circuitry, E_{mp} is the energy consumed by the transmitter amplifier and $d_0 = \sqrt{E_{fs}/E_{mp}}$ is the distance threshold between the transmitter and the receiver over which the multi-path fading channel model is used. The energy consumed to receive a message is $E_{rx}(sz) = sz \times E_{elec}$.

3.1. Classical clustering protocol

A classical clustering process is composed of two phases: the set-up phase and the steady-state phase, as depicted in Figure 3. When an event occurs in a random (uniformly-distributed) point of the network, nodes inside the event area wake up and start the clustering process. At the beginning of this phase, active nodes compete among each other to become a CH. Specifically, active nodes transmit their control packet to the sink node according to the specified random medium access protocol. In this chapter, CSMA control protocol is used, which is the default MAC protocol for the Mica platform. The control packet only comprises the node's ID and no data is transmitted at this point. The first node that successfully transmits this packet becomes the CH. All nodes involved in event-reporting immediately send their signaling message to the sink node. Therefore, the sink node selects the first node that successfully transmits a signaling message and then it broadcasts a signaling message over the network for a CH notification. Thus, the rest of the nodes become CMs. In the steady phase, CMs send their data in a scheduled fashion using a Time Division Multiple Access (TDMA) protocol. Note that the CH assigns slots to its CMs to accomplish the successful data transmission. Then, the CH aggregates the data values received from its CMs with its own, and sends the resulting data to the sink node.

3.2. Proposed compression clustering protocol

The proposed clustering Compression Cluster-based scheme in Spatially-Correlated Regions (CC_SCR), process is also composed of the same two phases, namely: the set-up phase and the steady-state phase. As in classical protocol, the set-up phase happens whenever an event occurs in a region of the network. However, in CC_SCR, the active nodes send their first measured data value to the sink node, i.e., they no longer send just their control packet. Instead, active nodes send a data packet. The reason for this is that this sensed data is used in CH selection procedure. Indeed, this entails an extra energy consumption at the set-up phase, compared to classical protocol. However, this first data transmission allows important energy savings in the steady-state phase.

It is important to note that CC_SCR is best suited for environments where event conditions are fairly stable during of the duration of the event. This is due to the fact that the CH is chosen according to the first sensed data. Hence, if event conditions suffer a high variation, the originally-selected CH may no longer render acceptable energy savings. An example of such an application is a fire surveillance forest, in which, when a fire occurs in a region, the temperature remains stationary for the duration of the fire in the region. Another example includes target-tracking. In this kind of application, the target is the source of the measured data at sensor nodes, such as light or temperature. Here, the measured data remains the same whenever the target stays in the same place and hence the sensor nodes sense the same measured data during the presence of the target. Next, we describe the set-up and the steady-state phases of the proposed algorithm.

- In the set-up phase, after receiving the first data packets from all the active nodes, the sink node calculates the difference between the data from node n_i and those from node n_j, $i \neq j$. Next, these differences are summed over. We call this sum of the difference between data values S_i. Then, the sink selects as CH the node which minimizes the total difference, calculated value S_i, between each node n_i and node n_j, $i \neq j$. Finally, the sink broadcasts a control message to the active nodes in order to notify the node selected as CH. Therefore, the rest of the nodes consider themselves as CMs. Note that there is no need for the CMs to send any extra packets since the sink already knows which nodes are active.

- In the steady-state phase, the CMs send the difference between their sensed data and the CH's data value, which corresponds to a compressed value, called Δ_i, rather than the complete data packet, $value\_CM_i$. Therefore, $\Delta_i = |value\_CM_i - value\_CH|$ represents the difference between the *i-th* CM's data value $value\_CM_i$, and the corresponding CH data value $value\_CH$. In order to perform this compression, the CH sends its complete sample data value to the CMs at the beginning of each event occurrence. Therefore, CMs send only the Δ_i to the CH. The main advantage of the proposed scheme is that the S_i calculation is made at the sink node, which is not energy or memory-constrained.

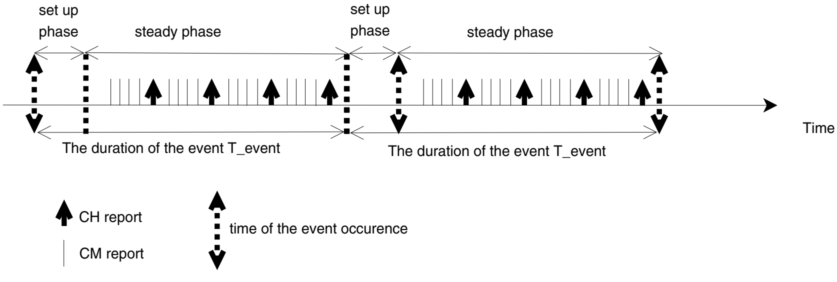
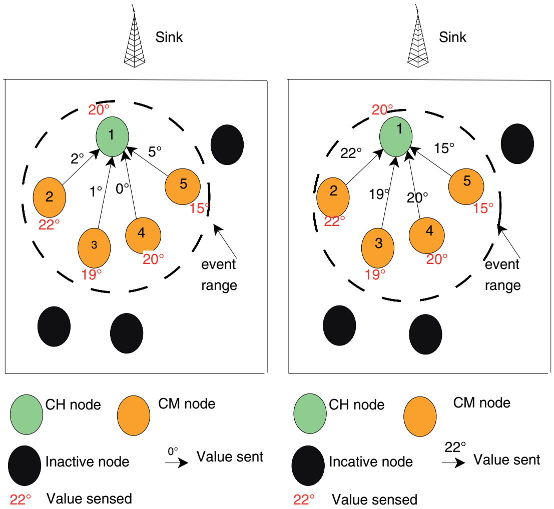


Figure 3. System operation.

3.2.1. Example

To illustrate the protocol's operation, let us consider the following example as presented in Figure 4(a) (Figure 4(b) shows the case of a classical scheme). We consider five active nodes n_1, n_2, n_3, n_4 and n_5 in the event region *e* which covers a region of range *R_event* as in Figure 1. In this example, we consider temperature as the sensed measurement value. Nodes n_1, n_2, n_3, n_4 and n_5 sense the value of 20°, 22°, 19°, 20° and 15°, respectively, and they send the values to the sink. When the sink receives the data values, it calculates the S_i value for each node n_i. The node which has the smallest S_i is considered as the CH.

The following calculation is done at the sink. For node n_1:
$|20 - 22| = 2$
$|20 - 19| = 1$
$|20 - 20| = 0$
$|20 - 15| = 5$
The sink node calculates $S_1 = |20 - 22| + |20 - 19| + |20 - 20| + |20 - 15| = 8$
For node n_2:
$|22 - 20| = 2$
$|22 - 19| = 3$
$|22 - 20| = 2$
$|22 - 15| = 7$
The sink node calculates $S_2 = |22 - 20| + |22 - 19| + |22 - 20| + |22 - 15| = 14$



(a) Proposed protocol. (b) Classical protocol.

Figure 4. Example of the system operation.

For node n_3:

$|19 - 20| = 1$

$|19 - 22| = 3$

$|19 - 20| = 1$

$|19 - 15| = 4$

The sink node calculates $S_3 = |19 - 20| + |19 - 22| + |19 - 20| + |19 - 15| = 9$
The calculation for node n_4 is the same as node n_1.
For node n_5:

$|15 - 20| = 5$

$|15 - 22| = 7$

$|15 - 19| = 4$

$|15 - 20| = 5$

The sink calculates $S_5 = |15 - 20| + |15 - 22| + |15 - 19| + |15 - 20| = 21$.

Finally, it selects either node n_4 or node n_1 as a CH. Both nodes minimize the total difference value measured. The other nodes become CMs. During the steady phase, CM nodes send their Δ_i value to the CH rather than their complete value. In this example, n_3 sends the sample value of 2 rather than the complete sample value of 22. Note that the compression data sent in our scheme involves sending less coded bits compared to a complete data that is sent in the classical scheme. Therefore, considerable energy saving is achieved in our scheme, as can be seen in Sections 4 and 5. Note that this election algorithm can be achieved at nodes in a distributed manner. Distributed scheme here means that nodes first exchange data and then the calculations specified at the sink node will run at the level of individual nodes. However, in this scheme nodes will receive a considerable amount of data, which may complicate the election process as nodes have a limited local memory.

4. Analytical results

In this section, the mathematical model of classical, single hop and CC_SCR protocols are described. The total energy consumed in the network, E_{total}, for the duration of an event, can be calculated as follows:

$$E_{total} = E_{competing} + E_{reporting} \tag{2}$$

where $E_{competing}$ is the energy consumed during the cluster formation phase and $E_{reporting}$ is the energy consumed during the steady-state phase. We calculate in the following $\mathbb{E}[E_{total}]$ as the expected energy consumed through the network for a single hop protocol and for both the classical and CC_SCR protocols.

$$\mathbb{E}[E_{total}] = \mathbb{E}[E_{competing}] + \mathbb{E}[E_{reporting}] \tag{3}$$

4.1. Classical protocol

For classical protocol, where no data compression is enabled, we first calculate $\mathbb{E}[E_{competing}]$. The energy consumed in the cluster formation phase is due to the signaling packet transmission of the active nodes in the event area directly to the sink plus the reception of the signaling packet from the sink to the active nodes, then:

$$\mathbb{E}[E_{competing}] = m \times [E_{tx}(S, R_t) + E_{rx}(S)] \tag{4}$$

where $m = N\pi R\_event^2/A^2$ is the expected number of active nodes in the range R_event when the disk is totally included in the area $A \times A$ and N is the total number of nodes in the network. $S = 24$ bits is the size of signaling message, and

- $m \times E_{tx}(S, R_t)$ is the energy consumed to send m competing messages to the sink.
- $m \times E_{rx}(S)$ is the energy consumed by the resulting competing messages sent from the sink through the network.

$$\begin{aligned}\mathbb{E}[E_{reporting}] =& Nb_r \times [E_{tx}(S, R_c) + (m-1) \times E_{rx}(S) + (m-1) \times E_{tx}(fixe, R_c) \\ &+ (m-1) \times E_{rx}(fixe) + E_{DA} \times fixe + E_{tx}(fixe, R_t)]\end{aligned} \tag{5}$$

where $fixe$ bits is the size of the full data packet, Nb_r is the number of packets sent during the steady phase, and

- $E_{tx}(S, R_c)$ is the energy consumed from a signaling message sent by the CH to its CMs in order to send their data.
- $(m-1) \times E_{rx}(S)$ is the energy consumed by CMs to receive the signaling message.
- $(m-1) \times E_{tx}(fixe, R_c)$ is the energy consumed by CMs to send data to the CH.
- $(m-1) \times E_{rx}(fixe)$ is the energy consumed by the CH to receive data sent by the CMs.

- $E_{DA} \times fixe$ is the energy consumed by the CH due to data aggregation.
- $E_{tx}(fixe, R_t)$ is the energy consumed by the CH to send the aggregated data to the sink.

In the simulation, we take $fixe = 32$ and $Nb_r = 29$.

4.2. Single hop protocol

In the single hop protocol, there is no energy consumed during the set-up phase. Nodes start sending data packets directly to the sink, then:

$$\mathbb{E}[E_{total}] = Nb_r \times m \times E_{tx}(fixe, R_t) \tag{6}$$

where

- Nb_r is the number of reports sent to the sink during T_event sec.
- m is the number of nodes involved in data reporting.
- $E_{tx}(fixe, R_t)$ is the energy required to send a full data packet to the sink.

The interest of analyzing this case is to have an insight into the benefits of clustering schemes in event-driven WSNs.

4.3. CC_SCR

We now consider the case where the CC_SCR strategy is enabled. It is to be noted that the proposed scheme, CC_SCR, behaves in the same manner in the cluster formation phase as classical protocol, with the important difference being that nodes transmit the data packet instead of the signaling packet, then:

$$\mathbb{E}[E_{competing}] = m \times [E_{tx}(fixe, R_t) + E_{rx}(S)] \tag{7}$$

where

- $m \times E_{tx}(fixe, R_t)$ is the energy consumed to send m data packets to the sink.
- $m \times E_{rx}(S)$ is the energy consumed by the resulting compete message sent from the sink through the network.

In the steady-state phase, energy consumption is found as follows:

$$\begin{aligned}\mathbb{E}[E_{reporting}] = & E_{tx}(fixe, R_c) + (m-1) \times E_{rx}(fixe) + Nb_r \times [E_{tx}(S, R_c) \\ & + (m-1) \times E_{rx}(S) + (m-1) \times E_{tx}(S + log_2(\mathbb{E}[\Delta_i]), R_c) \\ & + (m-1) \times E_{rx}(S + log_2(\mathbb{E}[\Delta_i])) + fixe \times E_{DA} + E_{tx}(fixe, R_t)]\end{aligned} \tag{8}$$

where

- $E_{tx}(fixe, R_c)$ is the energy consumed to send CH data packets to the CMs.

- $(m-1) \times E_{rx}(fixe)$ is the energy consumed by CMs in order to receive data sent by the CH.
- $E_{tx}(S, R_c)$ is the energy consumed from a signaling message sent by the CH to its CMs in order to send their data.
- $(m-1) \times E_{rx}(S)$ is the energy consumed by CMs in order to receive the signaling message.
- $(m-1) \times E_{tx}(S + log_2(E[\Delta_i]), R_c)$ is the energy consumed by the CMs to send the compressed data to the CH.
- $(m-1) \times E_{rx}(S + log_2(E[\Delta_i]))$ is the energy consumed by the CH to receive the compressed data from the CMs.
- $E_{DA} \times fixe$ is the energy consumed by the CH due to data aggregation.
- $E_{tx}(fixe, R_t)$ is the energy consumed by the CH to send the aggregated data to the sink.

where $\mathbb{E}[\Delta_i]$ is the average data packet size which corresponds to the difference between the CMs' data and the CH's data. It is worth noting that when considering an uniform node distribution with a large N, the node that minimizes the distance in the R_event region will be located at the center of R_event. Therefore, to calculate $\mathbb{E}[\Delta_i]$, let us first calculate the average distance between active nodes and the CH, $\mathbb{E}[d_{toCH}]$. We denote by $\mathcal{D}$ the disk of radius R_event, i.e. $\mathcal{D} = \{(x,y) \mid x^2 + y^2 \leq R\_event^2\}$. Since active nodes are uniformly distributed in $\mathcal{D}$, we have

$$\mathbb{E}[d_{toCH}] = \int\int_{\mathcal{D}} \sqrt{x^2+y^2}dxdy = \frac{1}{\pi R\_event^2}\int_{r=0}^{r=R\_event}\int_0^{2\pi} r^2 dr d\theta = \frac{2R\_event}{3}. \quad (9)$$

We then calculate $\mathbb{E}[\Delta_i]$, the average data difference between the data at the CM and the maximum value at the CH C, considering that the CH is located at the center of the cluster. Indeed, in a highly dense WSN, such the one considered in this work, it is reasonable to consider that the CH is located at the center of the cluster, i.e., very close to the event origin.

$$\mathbb{E}[\Delta_i] = |C - \frac{C}{(\mathbb{E}[d_{toCH}]+1)^{\alpha}}| = C \times |(1 - \frac{1}{(\frac{2R\_event}{3}+1)^{\alpha}})| \quad (10)$$

Note that the previous model is an approximation of reality, in which an ideal channel is considered, i.e., there is no consideration of packet loss. According to the previous models, Figure 5(a) and 5(b) show the average energy consumed in the network of $N = 1000$ for different values of R_t and R_c, respectively.

In Figure 5(a), we set $R_c = 100\ m$ and $Nb_r = 29$, and in Figure 5(b), we set $R_t = 150\ m$, $Nb_r = 29$. These results show that the CC_SCR strategy is suitable when R_t is less than around 180 meters and R_c is greater than around 50 meters.

Exceeding these thresholds makes the competing process of the proposed protocol very costly in energy due to the full data packet sent to the sink during the set-up phase. Remember that classical protocol only transmits a control packet in this phase. Therefore, CC_SCR has the highest energy consumption when the distance from the cluster to the sink becomes considerable.

In addition, Figure 5(a) demonstrates that the single hop scheme achieves the greatest energy consumption as 1) its transmissions depend directly on R_t, and 2) it sends the full data packets. Figure 5(b) shows a steady energy consumption for the single hop scheme as it does not use a CH to shorten its transmission range, therefore all nodes use a costly direct-transmission R_t, which dramatically decreases network lifetime.

Figure 5(c) shows the average energy consumed in the network when varying the Nb_r parameter. Here, R_t and R_c are set to $300m$ and $100m$, respectively. The results show that significative energy savings can be achieved when increasing the number of reports sent from the CMs to the CH. Building on from these observations, in Figure 5(c) we observe that the single hop scheme achieves the greatest energy consumption as its transmission depends directly on Nb_r.

The point of intersections, R_t_inter, R_c_inter, and Nb_r_inter of Figure 5(a), 5(b) and 5(c), respectively, are calculated as follows.

$$R_t\_inter = \sqrt[4]{\frac{fixe\times(m\times E_{elec}+E_{mp}\times R_c^4)-Nb_r\times(m-1)\times\Delta\times[2\times E_{elec}+E_{mp}\times R_c^4]-(S-fixe)\times E_{elec}}{(S-fixe)\times E_{mp}}} \tag{11}$$

$$R_c\_inter = \sqrt{\frac{m\times fixe\times E_{elec}-m\times(S-fixe)\times E_{elec}-m\times(S-fixe)\times E_{mp}\times R_c^4-2\times Nb_r\times(m-1)\times\Delta\times E_{elec}}{-fixe\times E_{fs}+Nb_r\times(m-1)\times\Delta\times E_{fs}}} \tag{12}$$

$$Nb_r\_inter = \frac{-m\times(S-fixe)\times E_{elec}-m\times(S-fixe)\times E_{mp}\times R_t^4+m\times fixe\times E_{elec}+fixe\times E_{mp}\times R_c^4}{2\times(m-1)\times\Delta\times E_{elec}+\Delta\times E_{mp}\times R_c^4} \tag{13}$$

where $\Delta = fixe - S - \log_2 \mathbb{E}[\Delta_i]$.

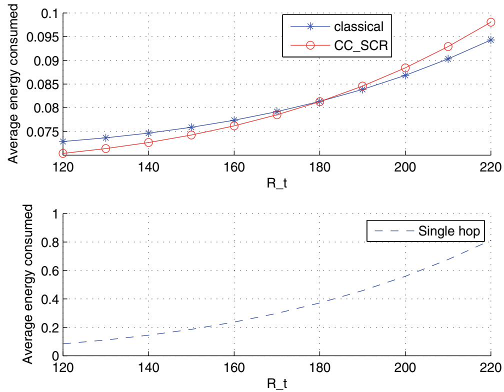
We conclude from Figure 5(a), 5(b) and 5(c) that 1) using a clustering scheme saves more energy than a single hop scheme, and 2) the application depends on R_t (referring to how far the sink node is from the area-sensed field), R_c (referring to the size of the area field), and Nb_r (referring to the number of data updating to the CH). More specifically, we conclude that for a fixed R_c, the higher R_t results a poor performance of CC_SCR concerning energy consumption. However, increasing R_c or Nb_r gives CC_SCR a better performance compared to the classical scheme. In addition, we conclude that both CC_SCR and the classical scheme outperforms the single hop scheme.

In the following, we analyze the network lifetime. Based on [28], the general definition of the average network lifetime, $lifetime$, can be expressed as follows:

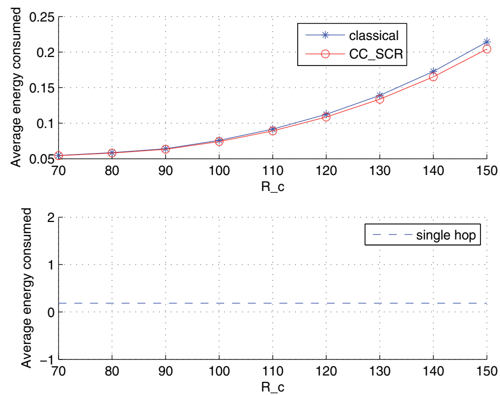
$$lifetime = \frac{\mathbb{E}[E_0\_total]}{\mathbb{E}[E_{total}]} \tag{14}$$

where $\mathbb{E}[E_0\_total] = m \times E_0$ is the average total residual energy in the area of R_event range, E_0 is the initial energy of a node and $\mathbb{E}[E_{total}]$ is the average energy consumed per unit of time (i.e., during T_event seconds).

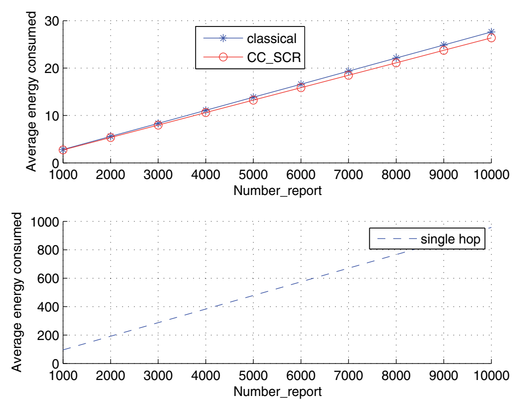
Using the same parameter settings as in the energy consumption analysis described above, Figure 6(a), Figure 6(b) and Figure 6(c) show network lifetime for different values of R_t,



(a) Varying R_t.



(b) Varying R_c.



(c) Varying Nb_r.

Figure 5. Analytical results of the energy consumption.

R_c and Nb_r, respectively. From Figure 6(a), we observe that CC_SCR outperforms the classical scheme when R_t is lower than around $180m$. The result of the single hop scheme, in Figure 6(a), shows that network lifetime decreases faster than in CC_SCR and in the classical schemes (Figure 7(a) shows the ratio gain of network lifetime).

From Figure 6(b), we observe that the more increase of R_t, the less network lifetime gain in CC_SCR, compared to the classical scheme. We also notice the steady network lifetime for the single hop scheme, but with a shorter value than for both the proposed and the classical schemes (Figure 7(b) shows the ratio gain of network lifetime).

From Figure 6(c), we observe that the higher the Nb_r, the longer the network lifetime, concerning the CC_SCR strategy compared to the classical scheme. We also observe that the network lifetime decreases faster in the single hop scheme, than compared to both CC_SCR and the classical schemes (Figure 7(c) shows the ratio gain of network lifetime).

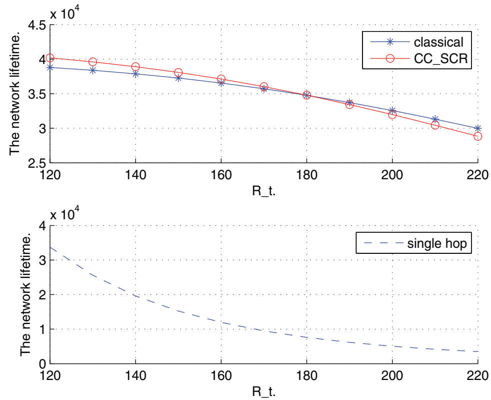
5. Simulation results

We use TinyOS [8] as a simulation tool. The parameters used for this set of results are presented in Table 1.

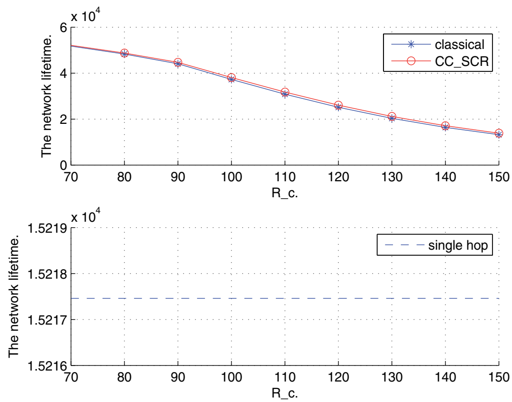
Parameter	**value**
E_{fs}	$10pJ/bit/m^2$
E_{mp}	$0.0013nJ/bit/m^4$
E_{elec}	$50nJ/bit$
E_{DA}	$5nJ/bit$
Signaling packet length S	$24bit$
Data value at the center of the event C	$250°$
Initial energy per node E_0	$10J$
T_event	200 *seconds*
R_event	60 m
R_c	100 m
R_t	400 m
Area A	$(100 \times 100)m^2$

Table 1. Simulation Parameters.

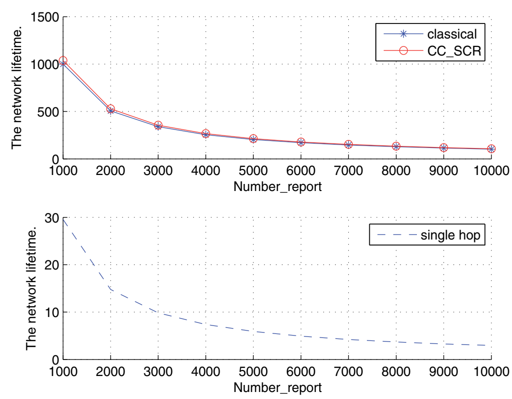
Figure 8 shows the average energy consumed in the network per unit of time for different concentrations of nodes. In this case, there are twenty simulated events. The results clearly demonstrate that our proposal outperforms the classical scheme. It can be seen that, as the number of nodes in the system increases, also the energy consumption increases. Indeed, when there is a high number of nodes in the network, there is also a high number of nodes that sense the event. Hence, the number of packet transmissions (both control and data packets) is much higher than in the case where just a few nodes are active per event. The main reason for the better performance of the proposed protocol is that only the difference, Δ_i, is transmitted rather than the complete data packet, during the steady-state. Note that this difference between classical and the proposed protocol increases concerning higher densities networks. The rationale behind this is that, in high density networks, nodes are closer to each



(a) Varying R_t.

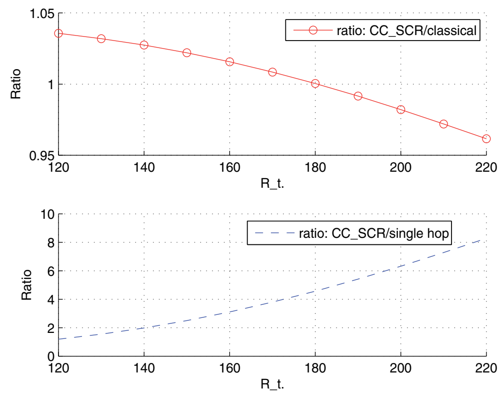


(b) Varying R_c.

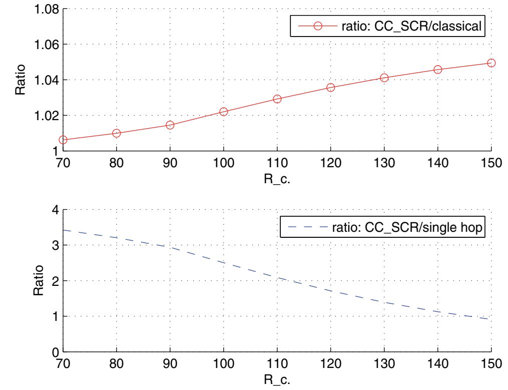


(c) Varying Nb_r.

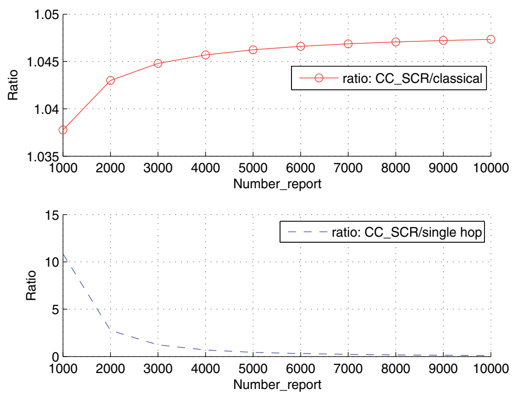
Figure 6. Analytical results of network lifetime.



(a) Varying R_t.



(b) Varying R_c.



(c) Varying Nb_r.

Figure 7. Ratio gain in network lifetime.

other, which in turn entails a higher correlation degree among the sensed data. This in turn renders a smaller packet size. Conversely, in the classical scheme, since the packet size is fixed, a higher density network only increases the number of packets transmitted, consuming a lot of energy.

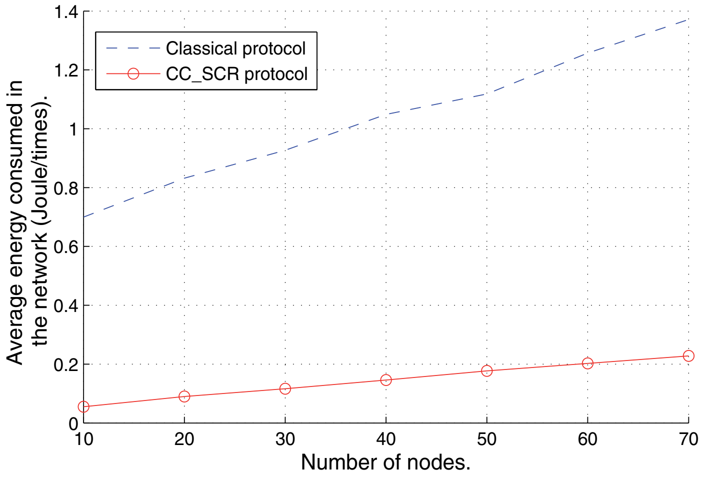


Figure 8. Average energy consumed per unit of time vs number of nodes.

Figure 9 shows the average energy consumed over time for 60 nodes in the classical, *CC_SCR* protocols, and also for one single hop to reach the sink. In order to explore the benefits of clustering architecture, a scenario where all nodes transmit directly to the sink is presented. For the proposed scheme, all active nodes transmit their initial packet to the sink in order to choose the reference node (note that in this case there is no CH). Then, the sink selects the node that minimizes the data difference, as explained in the previous section, and then transmits a control packet indicating the ID of the reference node. Then, for data transmission, active nodes only transmit their difference, Δ_i, directly to the sink. The results demonstrate clearly that *CC_SCR* conserves more energy compared to the classical scheme. Also, it is clear that the choice of clustering scheme offers more energy savings than the single hop scheme. The ratio gain presented in Figure 10 may reach up to 11 times more energy conservation than the classical scheme, and up to 119 times more energy conservation than the single hop scheme, which are considerable results.

Figure 11 shows the average energy consumed for different values of *R_event* region. When *R_event* is varied, also the number of active nodes per event is modified accordingly. Figure 12 shows the number of active nodes per event. It can be seen that the average number of active nodes for both the classical and proposed scheme is approximately the same. Indeed, the proposed mechanism has no impact on the number of active nodes. Note that, by increasing the number of active nodes, energy consumption also increases. Observe, for instance, that energy consumption, when $R\_event = 30$ meters, is less than the consumption when $R\_event = 60$ meters and 90 meters. In each scenario, we observe that, by enabling our compression scheme, energy consumption over the network is reduced, therefore extending network lifetime.

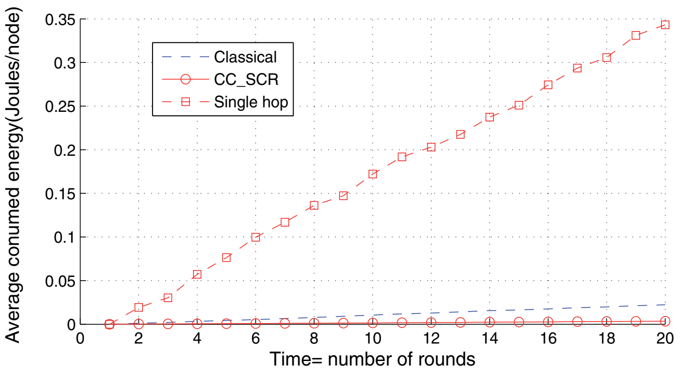


Figure 9. Average energy consumed over time.

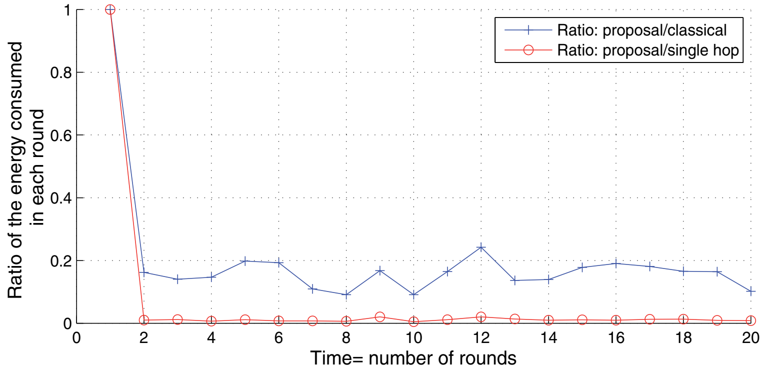


Figure 10. The ratio energy gain for each event occurrence.

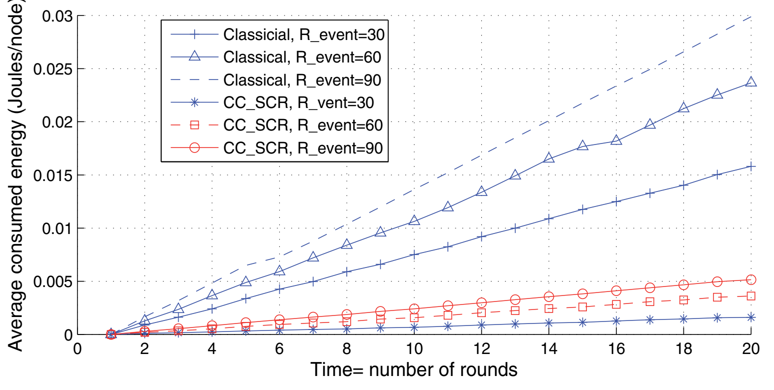


Figure 11. Average energy consumed while varying the *R_event* region.

Figure 13 shows the average energy consumed for different values of *T_event* period. Increasing *T_event* also increases the period of the steady-state phase and the number of data reported, thereby it can be seen as an increase in the energy consumption. That explains why the energy consumed by $T\_event = 200$ seconds is less than that of $T\_event = 300$ and 400 seconds. In each scenario, we observe that enabling our compression scheme reduces energy

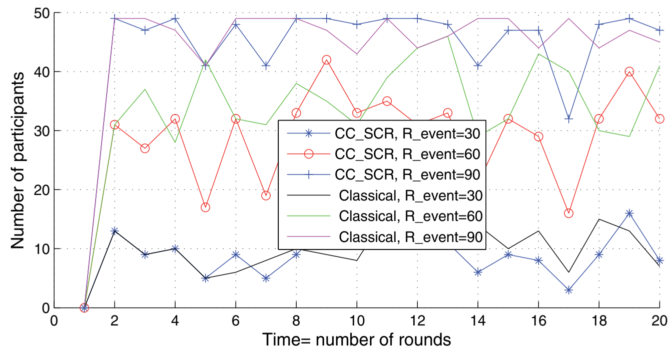


Figure 12. Number of active nodes for each event occurrence.

consumption over the network and thereby extends network lifetime. It is important to note that the proposed mechanism is particularly energy-efficient for high-event duration times. This is due to the fact that, as event duration increases, CMs in the classical scheme transmit many full-length packets while, in the proposed scheme, CMs also transmit many packets but with a much smaller length. This results in a slight increase in energy consumption for the proposed scheme while for the classical scheme there is a significant increase in energy consumption when the event duration increases.

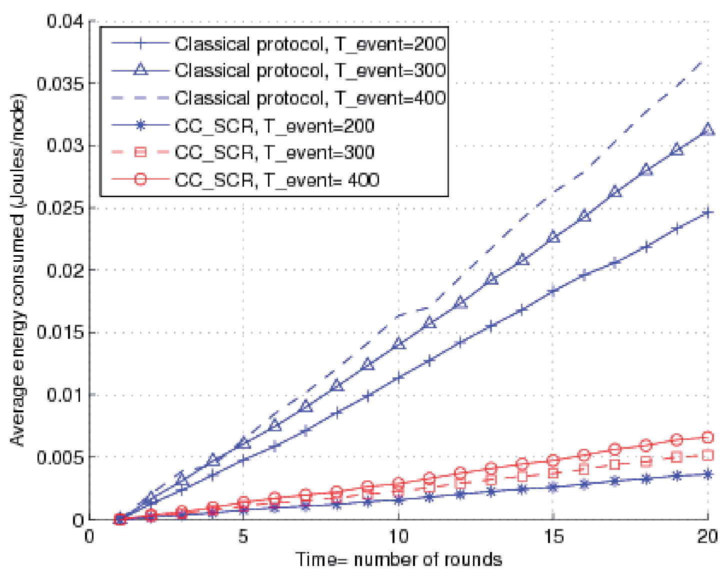


Figure 13. Average energy consumption vs number of rounds of length T_event.

Figure 14 shows the average energy consumption in CC_SCR when the aggregation technique is enabled at the CH, compared to the case where no aggregation is performed. The results clearly show that the aggregation technique conserves more energy (Figure 15 shows the ratio of the gain). The result is expected because when the CH aggregates the data of its CMs, the

CH only transmits one single packet to the sink, unlike the case when no aggregation is used, where the CH transmits each data value received from the CMs.

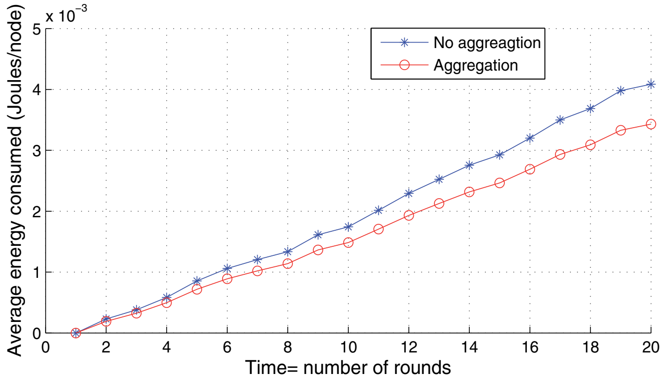


Figure 14. Energy consumed with aggregation and without aggregation.

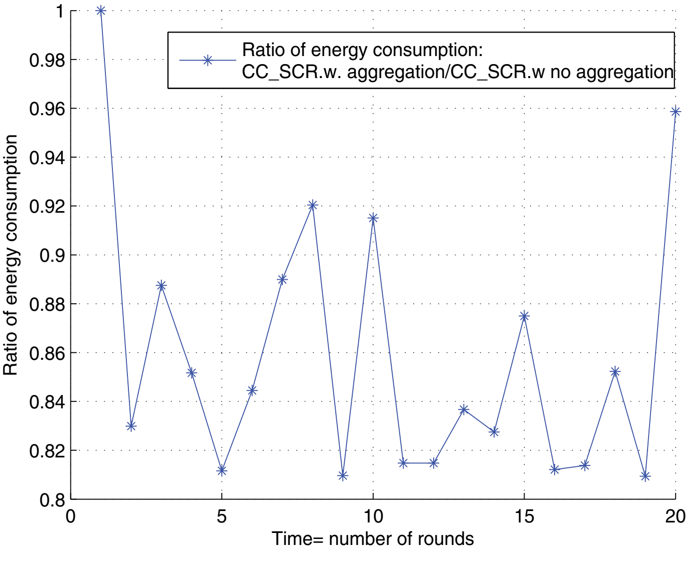


Figure 15. Ratio of energy consumed.

6. Conclusion and future work

In this chapter, we have proposed a novel clustering scheme, namely, *CC_SCR* protocol, which uses a compression technique for event-driven applications in WSNs. The proposed clustering scheme is based on selecting the node that reduces the packet size among all the active nodes in the system. The sink selects the node which minimizes the total amount of data as a CH, therefore increasing the efficiency of the compression technique by sending only the difference, rather than the complete data value, to the CH. To analyze the performance of the proposed scheme compared to both single hop (i.e., direct transmission to the sink) and classical schemes, an approximate mathematical model for energy consumption was developed. In addition, we implemented the *CC_SCR* in TinyOS, and for different system parameters, simulation results conclude that, considering the spatial correlation in the communication of WSNs, achieves significant energy conservation compared to a classical clustering scheme. The ratio benefit may be up to 11 times that of the classical scheme. As such, the proposed scheme greatly extends network lifetime. In future work, we aim to investigate a generalization of the clustering scheme in order to consider a higher number of events that can occur simultaneously in within the network, as it is the case in some environment monitoring applications. These results can be verified and deployed in our testbed. We also aim to include a mobility aspect to a certain number of nodes in the network, and consider this propriety in the clustering process in order to achieve considerable energy savings.

Author details

Sofiane Moad, Nizar Bouabdallah, Bruno Sericola and Yassine Hadjadj Aoul
IRISA/INRIA Rennes - Bretagne Atlantique, France

Mario E. Rivero-Angeles
National Polytechnique Institute (IPN), Mexico

7. References

[1] J.N. Al-Karaki and A.E. Kamal. Routing Techniques in Wireless Sensor Setworks: A Survey. IEEE Wireless Communications, 11(6):6–28, 2004.
[2] H. Karl and A. Willig. *Protocols and Architectures for Wireless Sensor Networks*. Wiley-Interscience, 2007.
[3] I. Dietrich and F. Dressler. On the Lifetime of Wireless Sensor Networks. ACM Transactions on Sensor Networks (TOSN), 5(1):1–39, 2009.
[4] N. Kimura and S. Latifi. A Survey on Data Compression in Wireless Sensor Networks. IEEE International Conference on Information Technology: Coding and Computing (ITCC), *Nevada, USA*, 2:8–13, April 2005.
[5] A.A. Abbasi and M. Younis. A Survey on Clustering Algorithms for Wireless Sensor Networks. Elsevier Computer Communications, 30(14-15):2826–2841, 2007.
[6] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A Survey on Sensor Networks. IEEE Communications Magazine, 40(8):102–114, 2002.
[7] S. Moad, M. Rivero, N. Bouabdallah, and R. Langar. *CC_SCR*: Compression Cluster-based scheme in a Spatial Correlated Region for Wireless Sensor Networks. International Conference on Communication (ICC), *Kyoto, Japan*, June 2011.
[8] P. Levis, N. Lee, M. Welsh, and D. Culler. TOSSIM: Accurate and Scalable Simulation of Entire TinyOS Applications. ACM International Conference on Embedded Networked Sensor Systems (SenSys), *California, USA*, pages 126–137, November 2003.
[9] K.C. Barr and K. Asanović. Energy-aware Lossless Data Compression. ACM Transactions on Computer Systems (TOCS), 24(3):250–291, 2006.

[10] C.M. Sadler and M. Martonosi. Data Compression Algorithms for Energy-Constrained Devices in Delay Tolerant Networks. ACM International Conference on Embedded Networked Sensor Systems (SenSys), *California, USA*, pages 265–278, November 2006.

[11] F. Marcelloni and M. Vecchio. Enabling Energy-Efficient and Lossy-aware Data Compression in Wireless Sensor Networks by Multi-Objective Evolutionary Optimization. Elsevier Information Sciences, 180(10):1924–1941, 2010.

[12] D. Petrovic, R.C. Shah, K. Ramchandran, and J. Rabaey. Data funneling: Routing with Aggregation and Compression for Wireless Sensor Networks. IEEE International Workshop on Sensor Network Protocols and Applications (SNPA), *Alaska, USA*, pages 156–162, May 2003.

[13] T. Arici, B. Gedik, Y. Altunbasak, and L. Liu. PINCO: A Pipelined in-Network Compression Scheme for Sata Collection in Wireless Sensor Networks. IEEE International Conference on Computer Communications and Networks (ICCCN), *Dallas, USA*, pages 539–544, October 2003.

[14] S.S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed Compression in a Dense Microsensor Network. IEEE Signal Processing Magazine, 19(2):51–60, 2002.

[15] D. Wei, S. Kaplan, and H.A. Chan. Energy Efficient Clustering Algorithms for Wireless Sensor Networks. IEEE Workshops of IEEE International Conference on Communications (ICC), *Bejing, China*, pages 236–240, May 2008.

[16] O. Younis, M. Krunz, and S. Ramasubramanian. Node Clustering in Wireless Sensor Networks: Recent Developments and Deployment Challenges. IEEE Network, 20(3):20–25, 2006.

[17] D. Amaxilatis, I. Chatzigiannakis, C. Koninis, and A. Pyrgelis. Component Based Clustering in Wireless Sensor Networks. ACM Computing Research Repository, 2011.

[18] O. Younis and S. Fahmy. HEED: a Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad hoc Sensor Networks. IEEE Transactions on Mobile Computing (TMC), pages 366–379, 2004.

[19] S. Lindsey and C.S. Raghavendra. PEGASIS: Power-Efficient Gathering in Sensor Information Systems. IEEE Aerospace Conference, *Montana, USA*, 3:3–1125, 2002.

[20] M. Ye, C. Li, G. Chen, and J. Wu. EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks. IEEE International Performance, Computing, and Communications Conference (IPCCC), *Louisiana, USA*, pages 535–540, April 2007.

[21] B. Kusy, C. Richter, W. Hu, M. Afanasyev, R. Jurdak, M. Brunig, D. Abbott, C. Huynh, and D. Ostry. Radio Diversity for Reliable Communication in WSNs. ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN), *Chicago, USA*, pages 270–281, April 2011.

[22] J. Faruque and A. Helmy. Rugged: Routing on Fingerprint Gradients in Sensor Networks. IEEE/ACS International Conference on Pervasive Services (ICPS), *Beirut, Lebanon*, pages 179–188, July 2004.

[23] M.C. Vuran and I.F. Akyildiz. Spatial Correlation-based Collaborative Medium Access Control in Wireless Sensor Networks. IEEE/ACM Transactions On Networking (TON), 14(2):316–329, 2006.

[24] S. Pattem, B. Krishnamachari, and R. Govindan. The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks. ACM/IEEE International

Symposium on Information Processing in Sensor Networks (IPSN), *California, USA*, pages 28–35, April 2004.

[25] G. F. Smoot et al. Description of COBE DMR. IEEE Astrophysical Journal Letters (ApJ), pages 360–685, 1990.

[26] A. Jindal and K. Psounis. Modelling Spatially-Correlated Sensor Network Data. IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON), *Santa Clara, Canada*, pages 162–171, October 2004.

[27] W.R. Heinzelman, A. Sinha, A. Wang, and A.P. Chandrakasan. Energy-Scalable Algorithms and Protocols for Wireless Microsensor Networks. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), *Istanbul, Turkey*, pages 3722–3725, June 2000.

[28] Y. Chen and Q. Zhao. On the Lifetime of Wireless Sensor Networks. IEEE Communications Letters, 9(11):976–978, 2005.

Section 3

Applications

Chapter 11

Monitoring Technologies in Mission-Critical Environment by Using Wireless Sensor Networks

Yuanyuan Zeng, Kai Xiang and Deshi Li

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48185

1. Introduction

Monitoring in mission-critical environment with the deployment of wireless sensor networks (WSNs) is one of the most widely-used application areas. Mission-critical environment implies monitoring tasks happened in locations that are difficult to get access to or easy to lead destroy of the network deployment: e.g., scenarios in hazard and emergency monitoring situations such as chemical pollution, fire and explosion, environment-oriented pollutions and land security (Poovendran, 2010). In the near future it can be expected that surveillance areas will be equipped with a range of smart sensors to provide parts of overall event detection and service. Sensors are expected to be programmed to report periodically and also when they detect a sensor input that exceeds a threshold. WSNs are well suited to monitor physical signals, allowing detection, localization, tracking and reacting toward the mission-critical environment.

We focus on monitoring related technologies for WSNs in mission-critical environment: 1) the connectivity issues in mission-critical monitoring and the solutions. 2) The data collection in mission-critical monitoring are addressed. The data collection requires technologies that can guarantee performance such as timeliness, reliability, scalability and energy efficiency. It can be provided by designed routing, link scheduling and even cross-layer mechanisms. 3) In mission-critical environment, dynamic network topology leads more difficulties in event detection. The related recent detection models and frameworks are addressed. 4) A new promising way for mission-critical monitoring is to utilize wireless sensors and actuators together. The monitoring related technologies about wireless sensor and actuator networks are then talked about.

2. Connectivity issues

Wireless sensor networks based monitoring and surveillance systems are physical and engineered systems that bridge the cyber-world of computing and communications with the physical world whose operations are monitored, coordinated and controlled. The research on connectivity of wireless sensor networks usually aims to utilize dense deployment that helps to improve reliability and extends longevity of the network lifetime. For general WSNs, energy efficiency is one of the major constraints in wireless sensor networks. Sensors are very limited in their processing, computing and communication capabilities as well as the storage and power supply. Moreover, due to the sheer number of sensor nodes and the potentially mission-critical environment, it easily leads to node failures and network partitions after the deployment of sensor nodes. So, how to maintain connectivity with recovery under the mission-critical environment is very important and challenging issue.

2.1. Challenges for connectivity in mission-critical environment

In mission-critical environment, the systems are usually built upon wireless sensor networks that need to provide timely and valid information about the field. As emergency events spread throughout the surveillance area, it is possible that the sensing devices will be easily disconnected from the network or indeed be destroyed. The network topology changes rapidly in the surveillance scenarios as emergency events spread. And then may lead to failure of the network system eventually. In this context, connectivity maintenance is very critical in order to ensure timely, accurate, and scalable monitoring.

The "Routing Hole" problem is a very important and well-studied problem in network connectivity, where messages get trapped in a "local minimum" that can be incurred by network partitions. Actually, the incidence of routing holes increases as network density diminishes and the success rate of the greedy algorithm drops very quickly with network density.

2.2. Related work on routing hole

Some existing "face routing" algorithms are developed to bypass routing holes using geo-routing algorithms. GPSR (Karp et al., 2000) recovers "hole" by using the "right-hand rule" to route data packets along the boundary of the hole, which combining greedy forwarding and perimeter routing on a planar graph that represents the same connectivity as the original communication network. Some incremental improvements are proposed based on it (Yu et al., 2000; Powell & Nikoletseas, 2007). GEAR (Yu et al., 2000) picks a next hop node that minimizes some cost value such as distance or energy and gradually forms a better path to the destination by sending multiple data packets. At this point a temporary rescue mode is used to escape the local minimum. But a major pitfall of face routing algorithm is that no practical planarization algorithm is known. An interesting approach BOUNDHOLE algorithm (Fang et al., 2006) is proposed to uses the TENT rule to discover local minimum nodes and then "bounds" the contour or the routing holes. But this algorithm has a high overhead.

In the mission-critical application situations, holes feature prominently and can be expected to grow in size rapidly as monitoring events spread, thus demanding solutions that are robust and low complexity for quick reactions.

2.3. Dealing with routing hole in mission-critical monitoring

Given a wireless sensor network deployed in a surveillance area under mission-critical environment. Each sensor can adjust its maximal transmission ranges to one of the *k* levels: R_0, R_1, …., $R_{k-1}=R_{max}$ by using different transmission power level from P_0, P_1, till $P_{k-1}=P_{max}$. Actually, most motes support to work under multiple power levels nowadays. Initially, all sensors work in P_0 for minimal energy under this power level. A sensor node is programmed to trigger and transmit data after it detects event locally. During data transmissions, if a sensor node cannot find a next hop that satisfies the routing metrics, a routing hole occurs. Here, an approach by increasing transmission power is utilized to jump over the "hole". It tries to find another node as the next hop by increasing the transmission power gradually until the maximal power.

Each sensor with *k* levels of power setting: $\{P_0, P_1, P_2, \ldots P_{k-1}\}$ can be in *k* levels of maximal transmission range as: $\{R_0, R_1, \ldots R_{k-1}\}$. A function is defined to find appropriate transmission power by increasing the power as follows:

$$P = P_{cur+\iota+1}, \iota = 1,\ 2,\ 3 \ldots k-1 \tag{1}$$

where, *cur* is the current number of transmission range level among *k* levels, ι is the count of unsuccessful delivery. When a sensor is switched to adaptation approach, it increases transmission power gradually in levels of setting, if cannot deliver the packet successfully. A node is eligible for power increase according to formula (1) until one of the following conditions is satisfied:

1. It finds a node in an eligible node as the next hop according to the routing metrics.
2. When $P=P_{max}$; In this case, it will try to find the relay according to routing metrics. Otherwise, no eligible relay is found.

The node that works in a larger transmission range could still be adapted to lower transmission power for energy efficiency, when it satisfies: the node is in a good connectivity with its neighbourhood that is larger then a predefined threshold. A function to find appropriate transmission range by decreasing transmission power is defined as:

$$P = P_{cur-\iota'}, \iota = 1,\ 2,\ 3 \ldots k-1 \tag{2}$$

where, *cur* is the current number of transmission power level among *k* levels. ι' is the count of decrement. A node is eligible for power decrease until:

1. The minimum power has been reached.
2. There are two consecutive power levels such that at the lower level does not met the required routing metrics, but at the higher power level does.

3. There are two consecutive power levels such that at the lower level the required neighbourhood connectivity threshold does not met, but at the higher power level the required does.

An example of the solution toward sensor data transmissions when in routing hole is illustrated as shown in Figure 1.

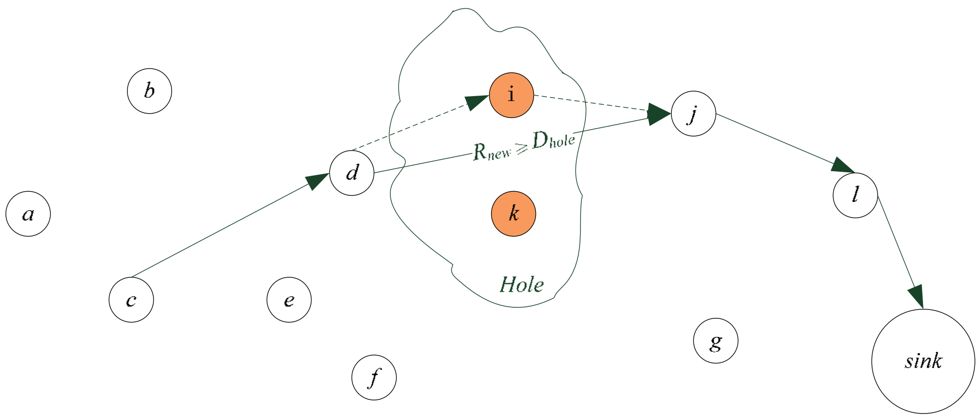


Figure 1. An illustration of the "Routing Hole Problem"

The node *d* can not find a suitable next hop and packets are stuck on node *i* and node *k*. So, node *d* is in local minimum. And we call that a "routing hole" occurs in the network. When we report "routing hole" on the hop of node *d*, the solution is to increase the transmission power and try to find the next hop with the higher power. As shown in Figure 1, node *d* increases its transmission power and find the next hop node. Actually, if the transmission range $R(R \leq R_{max})$ with the increased power $P(P \leq P_{max})$ exceeds the range of the "hole" (i.e., the adapted transmission range is larger than the maximal diameter of the hole), a new relay node *j* will be found located outside around the "hole". Then node *d* delivers the packets toward the new next hop *j* with higher power and jump over the "hole". In this way, the routing hole problem is solved by recover local connectivity by increased power.

3. Data collection technologies

The mission-critical monitoring systems are built upon wireless sensors that need to provide timely and accurate physical information. The rescue group relies on timely data update to be aware of on-time surveillance situations and reacts toward the emergency. In addition, as emergency events spread throughout the surveillance area, it is possible that the sensing devices will be easily disconnected from the network or indeed be destroyed. In this context, data collection requires a self-adaptive way in order to ensure timely, accurate, and scalable monitoring and surveillance. For a credible deployment of WSNs in mission-critical applications, the main properties of data collection that needs to be fulfilled are timeliness, self-adaptiveness, energy efficiency and scalability.

3.1. Challenges for data collection in mission-critical environment

The mission-critical surveillance applications are more challenging because of the time-varied connectivity and event scenarios. The network topology changes rapidly in the surveillance scenarios because of critical emergency events. The data packets have to be delivered effectively within the required time. The routes have to be adapted and reconfigured in order to continue operation for the network. The network can be in serious partitioned situations and not easy to be repaired, as emergency events destroy intensely in the surveillance area.

Power control schemes utilized in routing can help to increase the timeliness before the packet deadline especially for mission-critical situations. Increasing transmission power can effectively improve link quality and therefore reduce the number of transmissions to deliver a packet. The power increase reduces the throughput due to higher interference. Hence, supporting routing in mission-critical surveillance applications by power control alone is difficult.

3.2. Related work on application-specific communications

There are a lot of data collection related protocols and algorithms designed for ad hoc and sensor networks. Most of these schemes focus on energy efficiency and link node lifetime because of constrained node resources (Rogers et al., 2010).

Some sensor network system applications require real-time communication, typically for timely surveillance or tracking. The real-time routing design for general wireless sensor networks usually makes balance for real-time data delivery and energy consumption (He et al., 2003; Felemban et al., 2005; Chipara et al., 2006; Ahmed et al., 2008; Kim et al., 2009).

Beyond the above, the data collection design needs to consider application characteristics and requirements of the network system. Sivrikaya et al. proposed a stochastic routing mechanism for public safety applications such as emergency evacuations or rescue applications in wireless sensor networks based on Markov chains (Sivrikaya et al., 2009). Yang et al proposed a shortest path routing for mobile ad hoc and sensor networks by considering topology dynamics, i.e., the network topology changes due to energy conservation and node mobility (Yang et al., 2010). Li et al. proposed a solution to delay-bounded and emergency-efficient emergency event monitoring problem by an event detection model and a warning delivery model (Li et al., 2010). Li and Fen proposed a dynamic adaptive cooperative routing for emergency data in wireless sensor networks (Li & Fen, 2009). Byun et al developed a self-adaptive intelligent system for building energy saving and context-aware smart services (Byun & Park, 2011), where they proposed an energy-efficiency self-clustering sensor network and a node type indicator based routing protocol that considers the application requirement. Tseng et al. proposed a distributed 2D navigation algorithm to direct evacuees to an exit while helping them avoid hazardous areas (Tseng et. al., 2006). We have proposed a delay-sensitive routing scheme in wireless sensor networks for building fire monitoring (Zeng et al., 2011), which is adaptive to fire emergency scenarios such as node failure and dynamic network.

Beyond routing, being an essential part of data collection problems, media access control (MAC) protocols that handle with interference problems have received intense research attention. The proposed MAC layer schemes mainly include collision-based schemes such as CSMA/CA based schemes, and scheduling based schemes, as well as hybrid schemes. The collision-based schemes make collision-avoidance scheduling with node information of interference range usually two communication hops away. The scheduling based schemes are mostly based on TDMA slot allocation schemes. Sobral et al. proposed hybrid contention/TDMA-based (HCT) MAC, which is specially designed to work with ad-hoc wireless networks organized in clusters, providing timely bounded communications both inside and outside the clusters by resource reservation(Sobral et al.,2008).

There are some literatures with cross-layer design for data collection. The methods with joint optimal scheduling, routing and power control are proposed to achieve goals such as: fair rate allocation, minimize overhead and maximal resource utilization, etc. Recently, Lu et al. proposed to minimize average communication latency for the active flows with emergency efficiency by joint scheduling and routing in wireless sensor networks (Lu & Krishnamachari, 2007). Joseph et al. considered the problem of obtaining jointly optimal power control, routing and scheduling policies to ensure a fair utilization of network resources for energy harvesting sensor networks (Joseph et al., 2009).

3.3. Self-adaptive data collection framework

3.3.1. Intelligent data collection control model

An intelligent data collection control framework is used to circumvent the problem according to varied network scenarios. As illustrated in Fig. 2, the data collection control with adapted strategies is based on dynamic surveillance scenarios. The controller (sink) works as the gateway to control the data collection of the local sensor networks, which is used to collect physical data and interact with the Internet. During this, the controller gets the global knowledge of the emergency events occur. The controller updates and adapts the strategies of power, routing and scheduling in the network to achieve intelligent data collection.

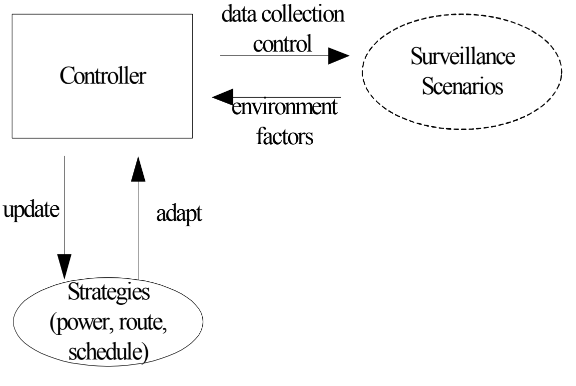


Figure 2. The intelligent data collection control framework

3.3.2. Mission-critical environment model

Considering characteristics of mission-critical surveillance, we use a finite-state discrete state set to model it. Usually, in most mission-critical surveillance, the scenarios of physical objectives could be in different finite states as: start, expand, diminish, etc. The set of finite discrete hazard states can be represented as $S=\{0,1,\ldots.K-1\}$. By partitioning the range of the received data delivery ratio of the controller into finite number of intervals, then physical objective model is constructed with finite hazard states. The number of set S is related with the partitions. The partitioning is executed by thresholds of the data delivery ratio. Let $\Gamma_0=0<\Gamma_1<\Gamma_2 \ldots<\Gamma_K=1$ be the thresholds of the data delivery ratio. The hazard scenario is in state k if the network data delivery ratio is between Γ_k and Γ_k+1.

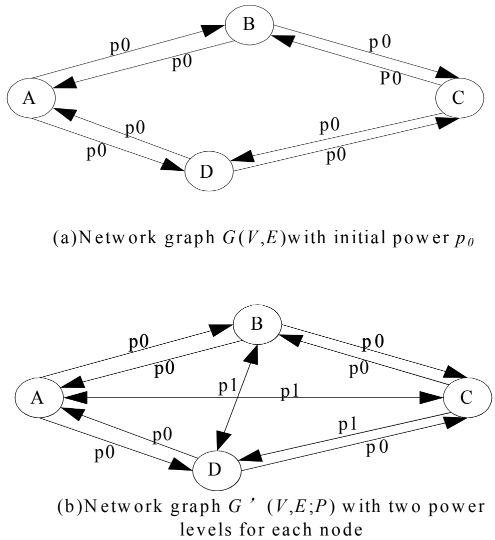
3.3.3. Elements of data control

The network system is formulated by a tuple $<S, A, R, \check{T}>$, where S is the discrete hazard state space. A is the discrete action space that is dependent on strategies taken. R: $S\times A\rightarrow \mathrm{R}$ is the cost function, which implies the quality of a state-action combination of the WSN-based network control system. $\check{T}$: $S\times A \rightarrow \Delta S$ is the state transition function, where ΔS is the probability distribution over state space S. The controller receives a performance-oriented signal from WSN in state $s_k\in S$, and selects a strategy $a_k\in A$ based on it.

For data collection control in mission-critical monitoring, the strategies include sets of selected power, route and schedule, i.e., a set of tuple pair (p_i, $node_i$, $slot_i$), where $p_i\in [p_{min}, p_{max}]$, $node_i\in [node_1, node_N]$ and $slot_i\in [slot_0, slot_{L-1}]$. Once the strategy is taken, the network system produces new performance signal according to the action. Then the controller receives the update cost $R\in\mathrm{R}$, which is used to evaluate the effectiveness of the control strategy. It is obvious that a higher power increases network connectivity and helps to find real-time delivery routes. In turn, it also increases the interference with larger interference ranges.

3.3.4. Search for strategy set

The search of strategy set for a WSN can use a multi-edge graph $G'(V,E';P)$ based on original network graph $G(V,E)$, which is shown in Fig. 3. For a connected network graph $G(V, E)$, each node can transmit data packets on each level of available transmission power. The each edge between u and v with $p_u\in [p_{min}, p_{max}]$ is drawn to form a multi-edge. It is well known that the derivation of all feasible scheduling in a single-channel wireless network is an NP-hard problem (Jain et al., 2003). Our problem is more difficult because of delay and interference control. Intuitionally, the number of possible concurrent transmission tuple set grows exponentially according to the increased number of allocations on directional links, power and slots. The search process is executed on the multi-graph G' and to find a "good" subset of transmission tuple series that satisfy delay and interference.



(a)Network graph $G(V,E)$with initial power p_0

(b)Network graph G' $(V,E;P)$ with two power levels for each node

Figure 3. The multi-edge graph G'

3.3.5. *Q-learning based data collection control*

The objective of the Q-learning based control algorithm is to learn the optimal transmission strategy (on power, route and schedule) that minimizes the cost function (i.e., gets maximal expected rewards). Let $Q_t(s_k,a_k)$ be the strategy quality. Essentially, the Q-value means the expected reward when taking strategy a under state s. The $Q(s,a)$ value are estimations of the optimal $Q^*(s, a)$, which implies the total discounted cost of taking strategy a in state s and then following the optimal control from then on. Accordingly, the optimal control strategy can be derived by selecting the set of power-relay-slot vector pair where the Q^*-value is minimized.

The learning procedure is to update the Q-value. When a packet transmission beginning at time t is finished at time t+1, the Q-value for state-strategy pair is updated by:

$$Q_{t+1}(s_t,a_t) = (1-\alpha_t)Q_t(s_t,a_t) + \alpha_t(R_t + \beta \min_{\alpha_{t+1}} Q_t(s_{t+1},a_{t+1})) \tag{3}$$

In Formula (3), α_t is the learning rate and in the range of (0, 1). β is the discount factor and is in the range of (0, 1) too. A constant learning factor is used so that the learning procedure can track the mission situations.

There is one data collection tree formed toward the sink during the transmission. Considering the multi-sink situation in the network, multi-agent Q-learning is utilized to find out the optimized strategy to improve system performance for all sink-oriented data collection trees. The motivation can be expressed as shown in Formula (4). Formula (5) shows the way to calculate the expectation of strategy cost. a is the selected strategy at state s. $\pi_i(a)$ is the probability for data collection tree toward sink i to select a specific strategy a, which achieves the equilibrium. Π_i is the set of strategies available. In this case, ($a_1 \ldots a_n$) is the joint action of trees under the equilibrium. $R_i(s, a)$ is tree i's total cost for the state given that agents follow the equilibrium strategies.

$$\min_{\pi^i \in \Pi^i} E[R^i(s,a)] \quad \textit{for any tree } i \tag{4}$$

$$E[R^i(s,a)] = \sum_i \pi^i(s,a) R^i(s,a) \tag{5}$$

The multi-sink data collection control process forms a stochastic non-cooperative game, i.e., multi-agent Q-learning process. Formula (4) shows the objective to achieve minimal cost for each data collection tree in the network. To achieve this, each tree finds a Nash equilibrium strategy when responses to the other trees for current state (Hu & Wellman, 2003). Let *NashQi* be the current agent's cost in current state for the selected equilibrium. According to Hu & Wellman (Hu & Wellman, 2003), we get Formula (6). m is the number for players in the game. In order to calculate the Nash equilibrium, each agent i needs to know the other agents' Q-values. Then, each agent observes the other agents' immediate rewards and actions. So, agent i can update its Q-value according to other agents' Q-values, as shown in Formula (7).

$$\begin{aligned} Q_{t+1}^i(s_t, a_t^1, \ldots, a_t^m) &= (1-\alpha_t) Q_t^i(s_t, a_t^1, \ldots, a_t^m) \\ &+ \alpha_t (R_i^t + \beta NashQ_t^i(s_{t+1}, a_{t+1}^1, \ldots, a_{t+1}^m)) \end{aligned} \tag{6}$$

$$NashQ_t^i(s,a) = \sum_1^m \pi^i(s,a) Q_t^i(s,a) \tag{7}$$

To minimize the system cost, the multi-agent Q-learning algorithm has to explore possible strategies randomly and greedily choose the "good" strategy. As shown in Formula (8), γ is a constant factor between 0 and 1. The learning policy satisfies the GLIE (Greedy in the Limit with Infinite Exploration) property.

$$\pi_{t+1}^i(a^*) = \begin{cases} \pi_t^i(a) + \gamma(1-\pi_t^i(a)), & \textit{if } a^* = \arg_a NashQ(s_t^i, a) \\ \gamma \pi_t^i(a), & \textit{if } a^* \neq \arg_a NashQ(s_t^i, a) \end{cases} . \tag{8}$$

In Algorithm 1: Line 1-6 is network initialization. In line 4, |A| is the number of control strategies. Line 7-12 is Q-learning procedure. In line 9, $R_1, \ldots R_m$ denote the rewards for tree 1 to tree m. $a_1 \ldots a_m$ denote the strategy taken by tree 1 to tree m. Line 10 is the Q-value update of each user according to Formula (7). The time complexity and space requirement of this learning algorithm is high when agent number is big. For 2-player Nash Q-learning, it has exponential worst-case time complexity. The space complexity is also exponential in the number of users.

Algorithm 1: Multi-agent Q-learning based Control

for i=1…m //m agents
  Let t=0, get the initial state s_0
  for all $s \in$S and $a \in$A
  $Q_0^i(s, a^1, ..., a^m) = 0, \pi_0^i(a) = 1/|A|$
  endfor
endfor
while (network execution condition is TRUE)
  Choose action a_i^t according to (8)
  Observe R1,…,Rm, a1,…, am
  Update Q_t^i for i=1…m according to (7)
  t=t+1
endwhile

3.4. Distributed cross-layer data collection

3.4.1. Network models

Given a WSN deployed in a surveillance area with n heterogeneous sensors and m controllers (i.e., sinks) that connected to Internet. Each sensor can adjust its maximal transmission ranges to one of the levels: r_0, r_1…r_{max} by using different transmission power levels from p_0, p_1, till p_{max}. Initially, all controllers work in default power p_0. A directed graph G (V, E) is used to model the network system. An end-to-end data delivery bound T_{max} is given as the maximum acceptable delay timeliness in reporting hazard event packets to the controller.

For each directed link e (u, v) in G, our data collection tries to find a data delivery route from data source s to sink d that satisfies the following:

- The found route delay from s to d is within the bound T_{max}.
- Within the power vector, each link on the path makes feasible schedule, i.e., interference-free scheduling.
- There exists a minimal end-to-end power allocation on each path link on this route:{p_0,p_1,p_2,….p_n} ,while $p_i \in$ [p_{min},….p_{max}], $i \in$ [0,1,….n].
- The data collection scheme is adaptive to the surveillance environment scenarios.

The physical model of interference is formed by signal to noise ration. Transmission from u can be successfully received by a receiver node v at slot t, if it satisfies:

$$\eta(p) = \frac{G_{uv} p_{uv}}{N_0 + \sum_{(x, y \in \tau \setminus \{(u,v)\})} G_{xv} p_{xy}} \geq \eta_{th} \tag{9}$$

In which, τ is the set of concurrent transmissions; p_{uv} is the power level set at the transmitter of node u for link (u, v). G_{uv} is the channel gain for (u, v) depending on path loss, channel fading and shadowing. η_{th} is a given threshold determined by QoS requirements such as bit error rate. N_0 is thermal noise power.

In protocol model, transmission methods are used to evaluate the interference impact. A node receiving from a neighbor, should be spatially separated from any other transmitter by at least a distance *D* defined by interference area. The "Interference Area" (IA) is defined as the maximal range that two concurrent transmissions would interfere with each other. For irregularity of radio, it is difficult to estimate IA by hops, because node *a* can reach node *b* doesn't mean node *b* can reach node *a*. In this case, "Possible Interference Area" (PIA) is defined as 2-hop neighborhood to estimate IA.

Assume that time is slotted into a non-overlapping equal time period called frame. The frame is divided into non-overlapping equal time periods calls time slots. Each time slot is a time slice enough for a data packet transmission and corresponding ACK. The frame structure is defined as shown in Fig. 4. Each frame includes three parts:

- Control part: A start beacon is broadcasted out and used for local time synchronization. It exchanges newly assigned slots of itself and its neighbors. In this case, each node exchanges the allocated slot of itself and its neighbors and then every node knows the allocated slots information among its 2-hop neighborhood. During this part, we use contention-based mechanism. Based on local allocated slots information in PIA, each node makes the interference-aware link slot allocation.
- Schedule part: The node schedules the assigned slot for the current link.
- ACK part: The node acknowledges the allocated slot in this part. If there is no flow on the allocated slot for continuous frames, the slot can be recycled. If interference occurs on allocated slot, then this slot is tagged as "interference slot". And then the node will choose the other available slots in control part in the next frame.

Each node has the same default frame length initially. In each frame, the default schedulable slot length is *L*.

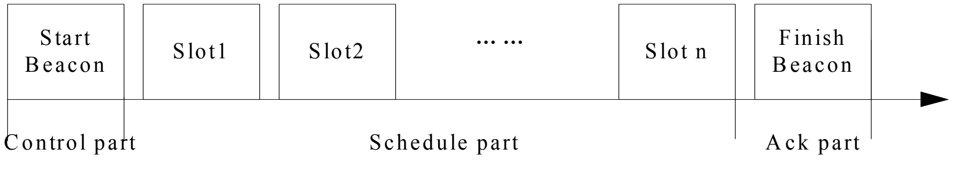


Figure 4. The frame structure

3.4.2. Cross-layer data collection schemes

The cross-layer can combine power control and schedule in data collection process and shift traffic towards the real-time delivery direction. Each node needs to maintain: 1) frame length and slot assignment information of it. 2) slot assignment information of its neighbors and 2-hop neighbors.

An admission control based method is used on each on-line link to make decisions on power allocation and schedule to achieve a real-time delivery with minor interference.

4. In "control" part: the node collects the slot allocation information. The per-link admission control is executed according to the delay deadline and interference constraints as follows:

- Calculate the available slots: the slots that are not being used by the other links among the node's possible interference area.
- Calculate the time left for end-to-end data delivery, which is denoted as slack. If satisfies *slack* – *t_sche* > 0, the link schedule is admitted. The link schedule time *t_sche* is calculated by: $t\_sche = t_c + t_s$, where t_c is the time waiting for the next link schedule. t_s is the time of scheduling slots for current link. The number of allocated schedule slots should satisfy the link flow.
- If the allocated slots are not being used for continuous number of frames, it can be recycled for other transmissions.

5. In "scheduling" part: the node tries to schedule the feasible slot for transmission, i.e., the real-time and interference-free slots are found with enough flow.
6. In "Ack" part: the node acknowledges the slot allocation among its 2-hop neighbors. In this part, the problems of slot reuse and the interference are dealt with.

- Slot reuse: if allocated slots are not used by certain number of continuous frames, they can be reused for other transmissions.
- Interference: if current link schedule interferences with other transmissions, the failure slots will be tagged as "unavailable" and then to choose other slots in the next frame.

The distributed cross-layer data collection is described in Algorithm 2. Line 1-21 describes the distributed cross-layer data collection algorithm, where the data delivery can combine the metrics of the existing real-time routing in WSNs. Line 22-34 is link admission control function. Line 1-5 shows that a node selects the next relay by metrics. If it finds a next hop satisfying admission control, then record the next hop with its assigned slot. Line 6-15 shows that the node will increase its power if cannot find a next hop satisfying successful admission control. The node increases the power by levels until it reaches the maximal power, and tries admission control each time. If the link admission control is successful with current power, then breaks out of the loop. Line 16-18 shows that the routing will be blocked, if the node fails to find a next hop satisfying admission control even with the maximal power. Line 23-28 shows how to assign slots for current link *e*. In line 24, $L \setminus slot(e')$ is the available slots except interference slots of current link. Line 27-28 is constraint condition that the time left for node data routing should be larger than the schedule time. Line 29-33 shows the return value of the function: if admission control is successful, the return value is 1. Otherwise, the return value is 0. It is obvious that algorithm 2 is with linear time complexity and suitable for mission-critical wireless sensor networks. For selection of *L* parameter, it tends to guarantee an interference-free transmission among local interference area. So, we can give an initialized *L* as $|\Delta|^2+1$, Δ is max degree in the network graph.

Algorithm 2: Distributed cross-layer data collection

```
For each node u selects the next relay with p_cur
 If reaches v
  If link_admission_control(e(u,v); p_cur)
      record allocated slots for e
      return
  else
      do
        increase p_cur to p_cur+1 to reach v′
        p_cur ← p_cur+1
          If link_admission_control(e (u,v′);p_cur)
             record allocated slots for e
             return
          endif
      while p_cur<=p_max
      enddo
      If  p_cur>p_max
        routing is blocked
      endif
  endif
 endif
endfor
int link_admission_control(e(u,v);p_i)
{
 slot(e) ← subset of (L\slot(e′)) with size Flow_e;
 // L is the schedulable slot number in a frame
 //e′ is the 2-hop neighborhood PIA links of e
  s.t.
  slack – t_sche > 0
  If slot(e)!=Φ
     return 1
  else
     return 0
  endif
}
```

3.4.3. Data collection with priorities

In multi-event sensor-based cyber-physical surveillance systems, there are different emergency levels for data collection. Two priorities for events can be defined in emergency missions as: "*high*" and "*ordinary*". For example, we define event priority as "*high*" priority, when it happens at nodes in dangerous conditions such as short of energy. The event as "*high*" priority can be defined as it happens on specific positions that in crowded venues or that is easy to collapse in surveillance area. The priority of data packets in hazard shows the different levels of importance for data delivery. The priority of event shows the different levels of importance and emergency for multi-event mission-critical data delivery. In this case, the data collection requests with high priority event always have more advantages on real-time data delivery when compared with an ordinary event.

According to surveillance scenarios, the node with high event priority could double its frame as shown in Fig. 5. If a node with high priority event data packets cannot find its next hop with interference-aware schedule, this node broadcasts a request to try to double its number of slots in the frame. The double frame request is broadcasted out in the control part of a frame. During the control part, if current node receives other requests from neighbors, only the one sent out earlier is admitted. The other requests are blocked. The node doubles its frame will notify its 2-hop neighbors. These 2-hop neighbors will update their frame length to keep local synchronization. For the synchronization among 2-hop neighbors and their neighbors, the double frame method can guarantee that there is at least one successful synchronization chance during every two previous frames. For a node in the network, it can easily reconnect to the network during continuous two timers of cycles among its neighborhood.

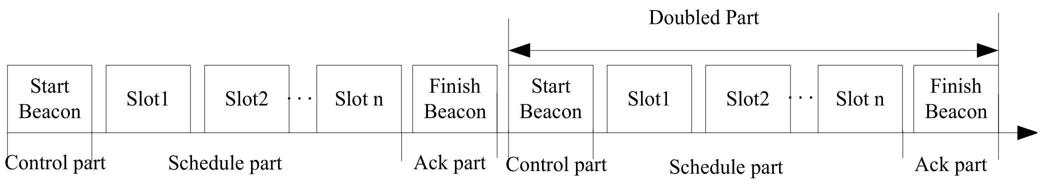


Figure 5. The structure of doubled frame

In Algorithm 3: Line 1-6 shows that each node selects the next hop until it cannot find the next hop with successful admission control. Line 7-15 shows the node increases its power level gradually to try to find an interference-aware next hop with a feasible schedule until reaches p_{max}. Line 16-23 shows if the node reaches p_{max} and cannot find a satisfying next hop with high event priority, then tries to double the frame length continuously to find a satisfying next hop. In lines 24-27, if it finds such a next hop with feasible schedule by double frame, the node records the slot information. Lines 34-39 show if no other frame double requests or current double request wins the tie, then both nodes on current link doubles its frame length. Otherwise, the function returns 0 to imply the failure of double request. Line 40-42 shows if current link nodes double the frame, they broadcast notification to 2-hop neighbors. Those nodes that receive such messages will update their frame length to keep synchronization.

The algorithm has linear time complexity, and easy to be performed in WSNs. The double frame scheme guarantees to deliver the high priority packets with high probability. A node with doubled slots can halve the frame till its default frame length and provide more chances for other data packets when necessary.

Algorithm 3: Distributed priority-based data collection

```
For each node u selects the next relay with p_cur
 If can reach v with p_cur
  If link_admission_control(e(u,v);p_cur)
    record assigned slots for e
    return
  else
    do
      increase p_cur to p_cur +1 to reach v'
      p_cur ← p_cur +1
      If link_admission_control(e (u,v'); p_cur)
          record assigned slots for e
          return
     endif
    while p_cur <=p_max
    enddo
    If  p_cur >p_max && event_priority== "high"
            do
                 p_cur ← p_max
                 If  !double_frame(e (u, v'))
                        break
                 endif
            while !link_admission_control (e(u, v'); p_cur)
             enddo
              If link_admission_control(e(u,v); p_cur)
                 record assigned slots for e
                 return
          endif
    endif
  endif
 endif
endfor
int double_frame(e(u,v))
{
  If no other requests or current request wins the tie
       Frame_ length(u)= Frame_length(u)*2
       Frame_length(v) = Frame_length(v)*2
  else
        return 0
  endif
  updates frame length among 2-hop neighbors
  return 1
}
```

3.4.4. Comparisons and simulation results

We verify the above data collection schemes for sensor based cyber-physical system in mission-critical hazard surveillance applications by simulations using ns2 network simulator. We consider the building fire hazard as a case study. To generalize and simplify the surveillance blueprint, we use a grid topology. The simulation results can represent performance for a general case. We choose 100 nodes that are distributed as grid topology in a 100m×100m area (Zeng et al., 2011). Each node can work under 3 power levels. We place 1-4 sinks on the corner of the simulation areas, respectively. We simulate and test our algorithms with an epidemic building fire model. Within the simulated area, a fire start point breaks out randomly 30s after the simulation is started. We define "high" priority hazard event when the event source node is with less than 10% energy. We use some metrics for performance evaluations. End-to-end delay is the total time needed for a data packet sent out till it is received correctly by a sink node. The packet delivery miss ratio (use "miss ratio" in the following paragraph) is the ratio of all packets missed because of the delay bound and interference versus the total packets sent out. The energy efficiency is evaluated by average residual energy in the network.

Fig. 6 shows the Q-learning performance as delay bound increases. As we increase the delay bound, the average residual energy increases too. More low power feasible transmissions available can decrease the energy consumption in the network when we relax the delay bound. In 2-agent learning case, two sinks are selected on the diagonal corner of the area. The source node reports data to the sink periodically until it fails. We set the end-to-end delay bound as 60ms and test the energy efficiency when compares 1-agent Q-learning and 2-agent Q-learning performance. Fig. 7 shows the average residual energy of 1-agent Q-learning and 2-agent Q-learning. From the result, we can observe that performance of 2-agent Q-learning is better than 1-agent Q-learning case because of one more sink. Fig. 8 shows the end-to-end delay as delay bound increases from 15ms to 100ms. The priority-based algorithm has slightly less delay, because it increases the delivery probability of the critical event in fire. Fig.9 shows the miss ratio performance. Fig. 10 shows the node average residual energy. We consider the high priority of critical event occurred by low-energy nodes in our simulations, those nodes deliver data packets with high priority and then achieves better energy efficiency in the network.

We then compare our distributed algorithms with RPAR (Chipara et al., 2006) and EAR (Zeng et al., 2011). Fig. 11 shows the end-to-end delay results comparison. We observe that our two distributed algorithms achieve much better performance compared with the other two mechanisms. Because we consider not only real-time delivery but also the interference-aware transmission, more packets are delivered within the delay bound successfully. Fig. 12 shows the miss ratio of the four mechanisms. Fig. 13 shows average residual energy.

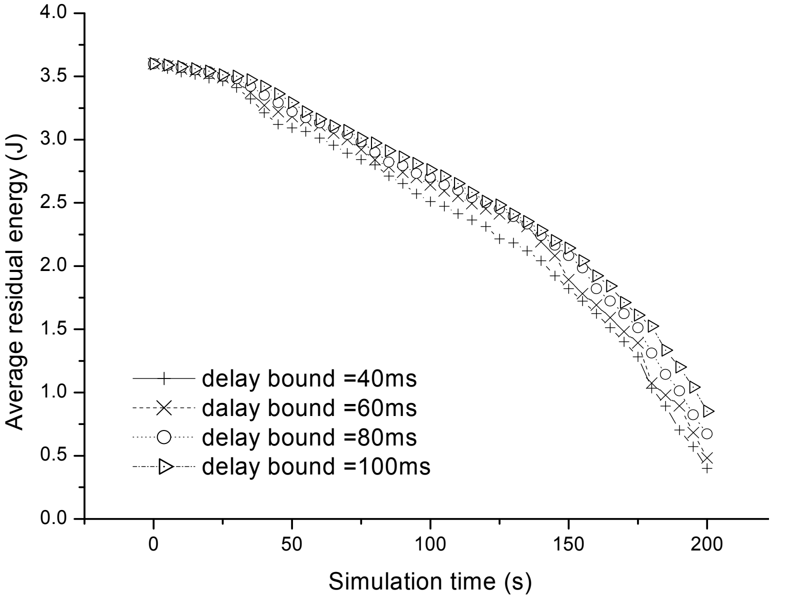


Figure 6. Energy efficiency by Q-learning as delay bound increases.

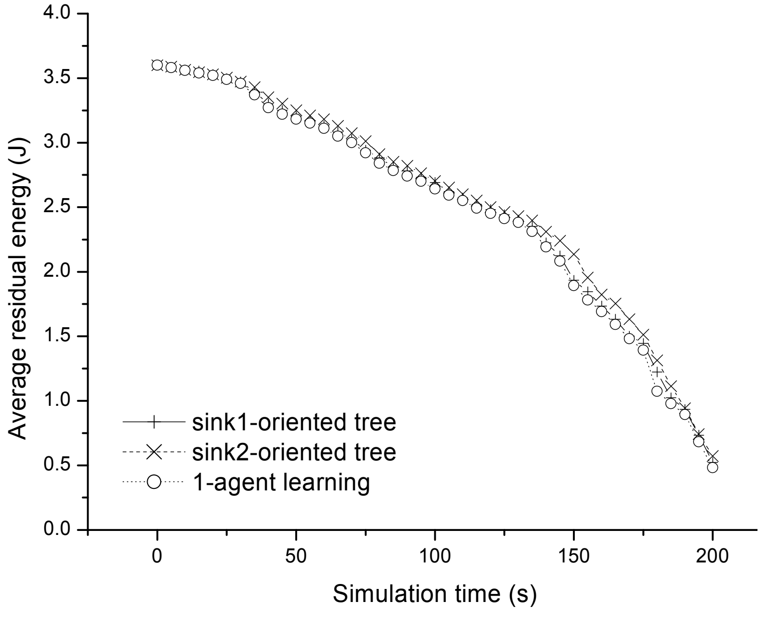


Figure 7. Energy efficiency of 2-agent Q-learning (delay bound=60ms).

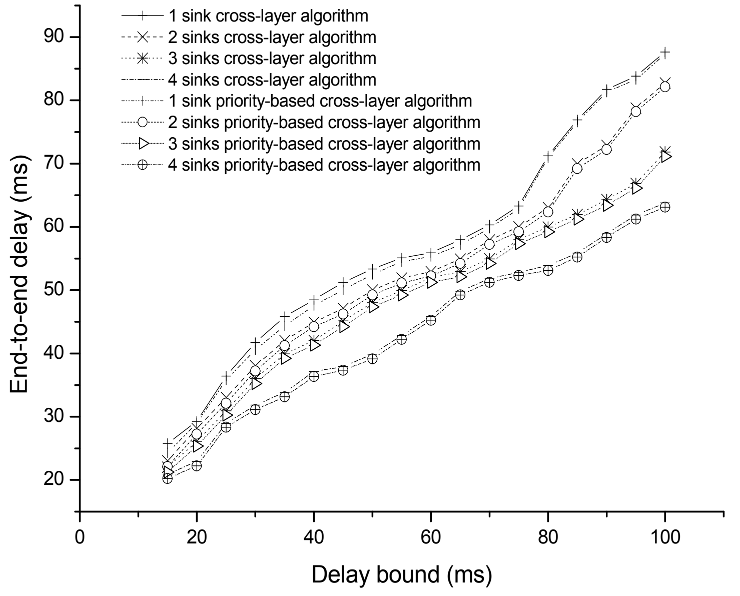


Figure 8. End-to-end delay as delay bound increases

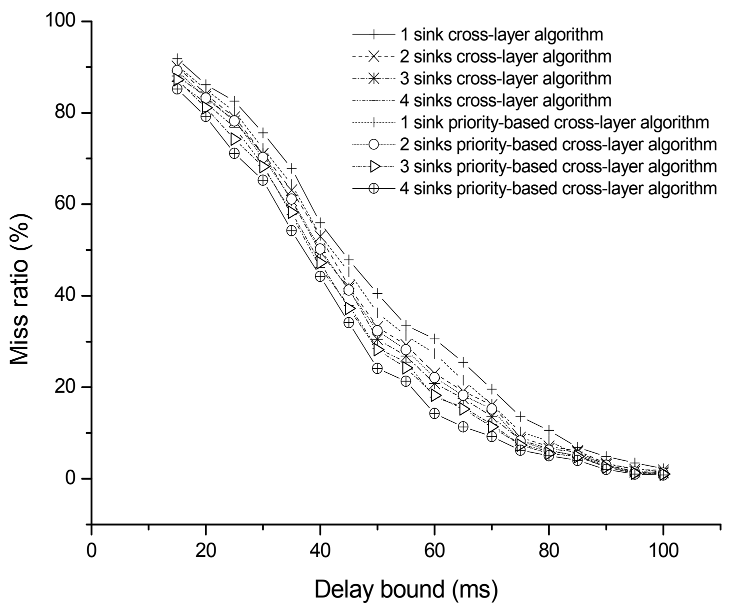


Figure 9. Miss ratio as delay bound increases

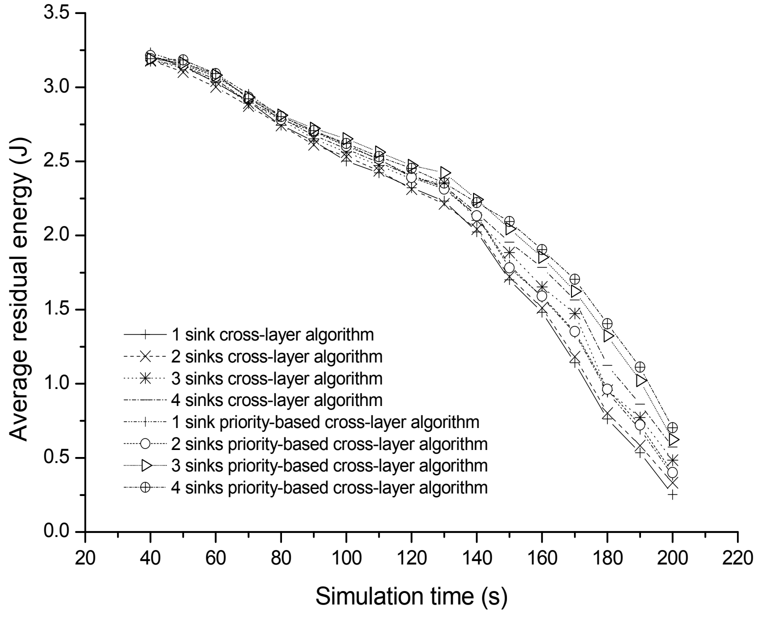


Figure 10. Energency efficiency with/without priorities (bound =60ms).

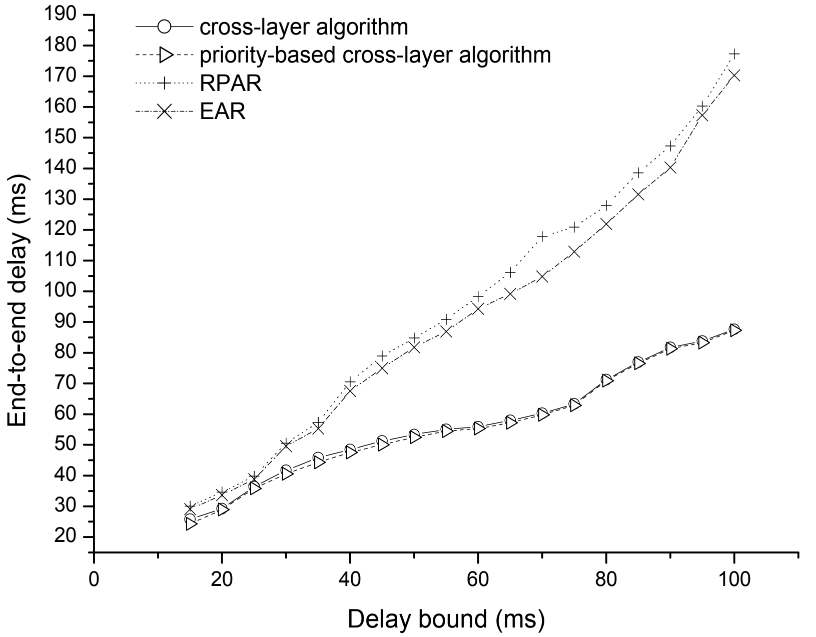


Figure 11. Comparison of end-to-end delay

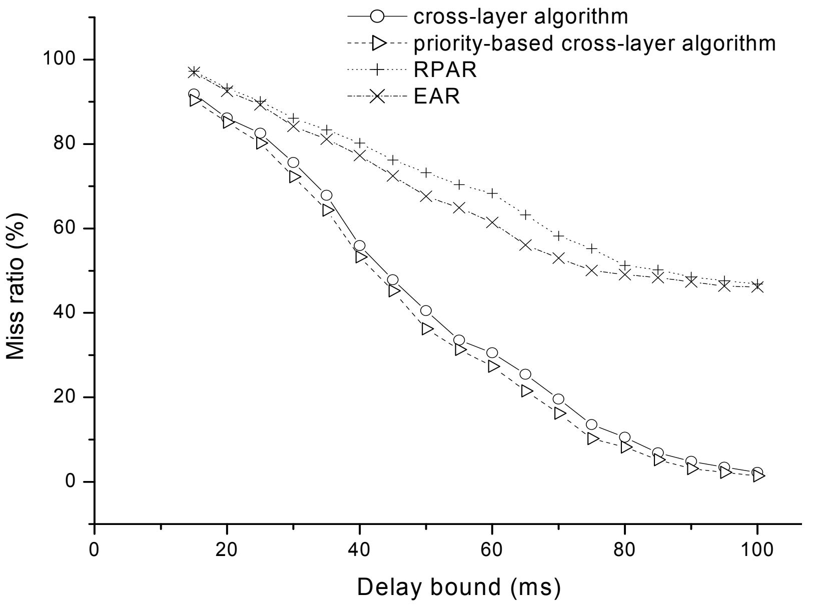


Figure 12. Comparison of Miss ratio

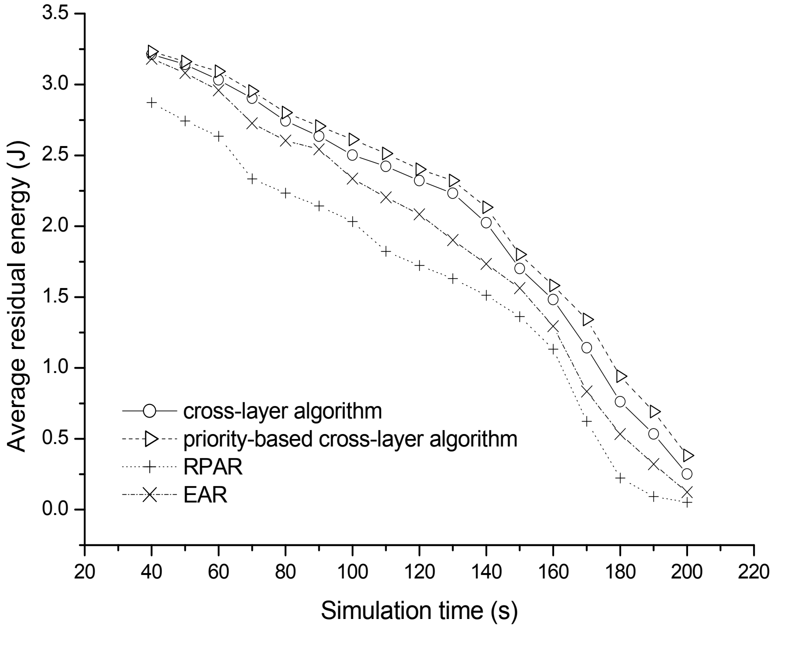


Figure 13. Comparison of energy efficiency.

4. Event detection technologies

Wireless sensor networks can be utilized continuously to monitor the physical phenomenon with the deployment in surveillance areas. While the sensor only reports its local information and data that may be uncertain and unreliable, effective processing, collaboration and analysis of data streams become necessary and important. According the characteristic of mission-critical application-specific environment, the design of detection approaches is facing more challenges. The detection models and architectures with good performance are needed that can effectively detect the event from data sensed within the timeliness and accuracy requirements.

4.1. Challenges for mission-critical detection

To achieve the goal of event detection especially emergency detection, we require the continuous assessment of the state of the surveillance area, forecasting the possibility of events. The volume of data that is gathered form the monitoring network usually can be enormous, particularly with high bandwidth devices such as video cameras. According to this large amount of sensing data, an efficient model that can process well the data is needed. In a majority of cases the information is insufficient and the inter-communications between different sensors detecting the emergency event are inadequate, which leads to detection inaccuracy and a potentially huge loss of life and resources. This is even worse for the complex and varied environment.

4.2. Related work on WSN event detection

A lot of work has been done for event detection in wireless sensor networks. The work can be categorized into different categories according to the approaches.

Firstly, sensing coverage is an important problem in WSNs and surveillance applications. One individual sensor is easy to get weak and noisy signal. The multi-sensor information fusion and aggregation algorithms have been proposed to make event detection by using the coverage of sensors. In this case, combining the multiple sensor capability to provide the enough detecting reliability by increased coverage obtained from multiple sensors. The basic objectives of this category approaches are to minimize the probabilities of detection errors while meeting the specific constrains of communication cost and energy efficiency. In which, energy saving is usually achieved by ensuring at least k nodes are awake to cover the detection location, leaving the other nodes asleep (Xing et al., 2005).

Some model-based approaches are proposed by using statistical analysis. Zhuang et al. proposed to use a weighted moving average based approach to remove outlier sensor data as noise (Zhuang et al., 2007). Gupchup et al. proposed to use well-established Principal Component Analysis to build a compact model of the observed phenomena (Gupchup et al., 2009). The Principal Component Analysis model is used to determine a single or a sequence of measurements that are dissimilar to the normal behaviour of the system. There are some solutions using a probabilistic noise distribution to mathematically model the sensing area and detect events (Rachlin et al., 2005).

Other solutions utilize machine learning based approaches to provide event detection. This category of approaches makes feature classification and exaction from sensing data (Benbasat & Paradiso, 2007; Eriksson et al., 2008; Kang et al., 2008). There are some approaches that propose to utilize Hidden Markov Models to deduce events (Singh et al., 2008).

Event detection by using WSNs is also application-specific. Especially for mission-critical applications, they need to impose stringent requirements for event detection accuracy and network lifetime, etc. Keally et al. proposed a confident event detection approach in WSNs that can adapt the detection capability with the runtime observations to save energy (Keally et al., 2010). There is some related work that combines activity recognition into sensor based event detection in body networks (Keally et al., 2011).

4.3. Detection cases for mission-critical monitoring

4.3.1. An integrated fire emergency response system: FireGrid

The FireGrid system (Upadhyay et al., 2009) aims to leverage a number of modern technologies such as wireless sensor networks and grid computing to aid building fire emergency response. As shown in Fig.14, to detect the building fire and make responses toward hazard events, there are different visions with different technologies in the system to communicate among fire steward, remote management, and collaborate with smart devices in the building such as sprinklers, evacuation indicators, etc. The building modelling is achieved by blueprints, maps and scenarios. The emergency response is dependent on the physical information that is achieved by sensor networks through sensing, transmission and then analysis and decision making upon the data. The decision making of responders is based on a knowledge based system and planning by remote experts, which makes the responses by sensor data processing and super-real-time simulations.

4.3.2. The CodeBlue infrastructure

The CodeBlue infrastructure (Lorincz et al., 2004) is an ad hoc sensor network infrastructure for emergency medical care-tracking the patients and first response application system, which is introduced by Harvard University in collaboration with various medical facilities. CodeBlue is designed to provide routing, naming, discovery, and security for wireless medical sensors, PDAs, PCs, and other devices that may be used to monitor and treat patients in a range of medical settings. The goal is to enhance first-responders' ability to access patients on scene, ensure seamless transfer of data among caregivers and facilitate efficient allocation of hospital resources. Fig.15 shows the CodeBlue system architecture. The wearable wireless sensors are put on patients' body to collect heart rate, blood oxygen saturation, and hearts electrical activity, etc., continuously. The sensing data can be displayed in real time and integrated into system record. The smart devices are programmed to alert medical personnel when vital signs fall outside of the normal conditions. Any adverse change in patient status can then be signalled to a nearby emergency terminal or paramedic.

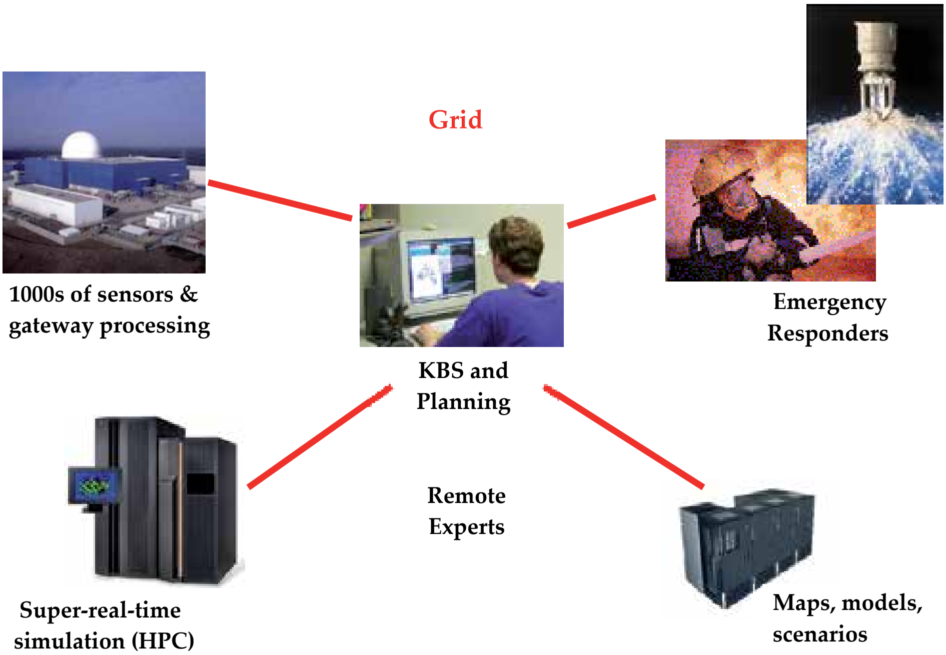


Figure 14. The FireGrid visions

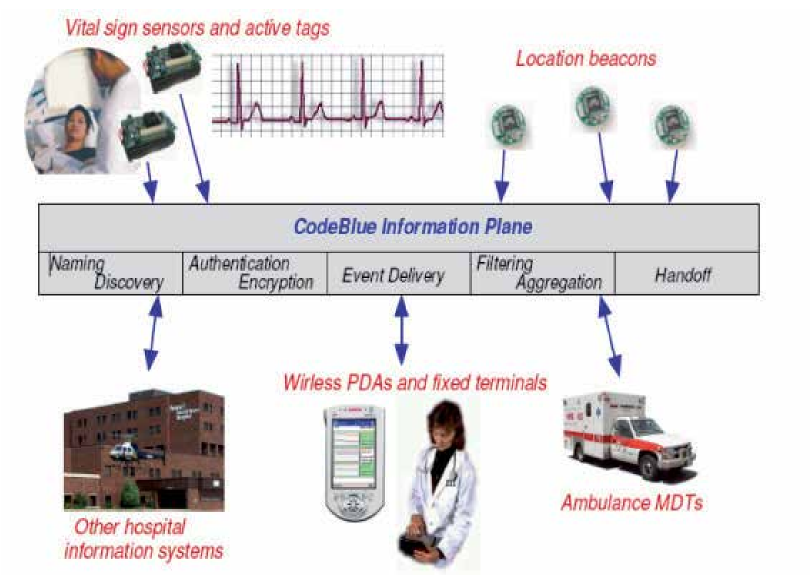


Figure 15. The CodeBlue system architecture

4.3.3. The NEMBES facilities management

For NEMBES project on Facilities management within buildings, the objective is to facilitate a multi-user environment precise data aggregation, acquisition and interpretation that can provide successful building automation (Zeng et al., 2009). We take building fire as a case study. We then design a real-time simulator for detecting and handling building fire emergency scenarios. The goals of this simulator are to provide for: 1) a dynamic virtual test-bed for population routing and networking algorithms during emergencies, 2) identification of building features that impact on evacuation scenarios, such as corridors prone to congestion, 3) visualising real-world emergency situations and predicting outcomes to inform rescue personnel as to the best rescue strategy or possible danger areas.

The underlying world model for this simulation is an object-based 2.5 dimension "building". Each floor of the building is a 2D collection of world objects, with the floors arranged in a spacial collection (ground floor, first floor, second floor etc). Stairs, fire escapes and elevators provide a mechanism for agents to travel between floors. This 2-and-a-half dimension model was chosen as it simplifies agent behaviour computations and allows for very clear visualisation of the emergency as it unfolds. The underlying building objects have analogues within the Industry Foundation Classes building model objects, such as walls, doors and so on. The simulation features multiple agents with dynamic behaviours navigating a building during an emergency. These agents are driven by a Sense->Plan->Act cycle and have basic memory. The two main classes of Agent are "Occupant" agents (persons present in the building, primarily driven by environmental cues such as direction signs or following crowds) and "Firefighter" agents (primarily driven by individual instructions, such as radio contact or personal "compass" direction). Agents will have steering and crowding mechanisms to accurately reflect real-life population movement. The underlying physical model of the world combined with such measures will provide useful knowledge as to areas in the building with excessive traffic and poor movement flow, or parts of a building which are of high-importance for evacuation (e.g. a main corridor).Fig. 16 shows a screenshot of our simulation for building fire.

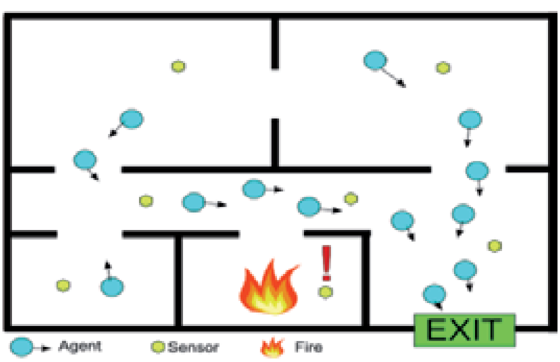


Figure 16. The NEMBES building fire simulator

4.3.4. Mobile sensing architectures

By using the mobility of people with cell phones, bicycles, cars, there are many platform and testbed projects are proposed based on mobile sensing networks.

BikeNet (Eisenman et al., 2007) is a platform of Mobile Sensing by using bicycles. In BikeNet, the real-time social networking is built among the cycling community. The bike area networking is a multifaceted sensing system that explores personal, bicycle and environmental sensing based on smart sensor motes and sensor-enabled Nokia N80 mobile phones. The mobile sensing data is collected through real-time and delay-tolerant communications to a networked repository, which is used to infer and find routes with low CO2 levels. There is provision of a web portal for cycling community that can access and share the real-time data of best routes with better air quality. Fig. 17 shows the BikeNet network architecture.

CarTel (Hull et al., 2006) is a distributed vehicle sensor network platform proposed by MIT Computer Science and Artificial Intelligence laboratory, which combines sensing, processing, analysis and visualization. CarTel platform is involved in several applications including commute and traffic portal, traffic mitigation using new predictive delay models, Pothole Patrol, fleet testbed and Wi-Fi monitoring, cars as mules etc. CarTel provides a simple query-oriented programming interface, handles large amounts of heterogeneous data from sensors, and handles intermittent and variable network connectivity. CarTel nodes rely primarily on opportunistic wireless (e.g., Wi-Fi, Bluetooth) connectivity—to the Internet, or to “data mules” such as other CarTel nodes, mobile phone flash memories, or USB keys—to communicate with the portal. Fig. 18 shows CarTel in traffic delay estimation application (Thiagarajan et al., 2009).

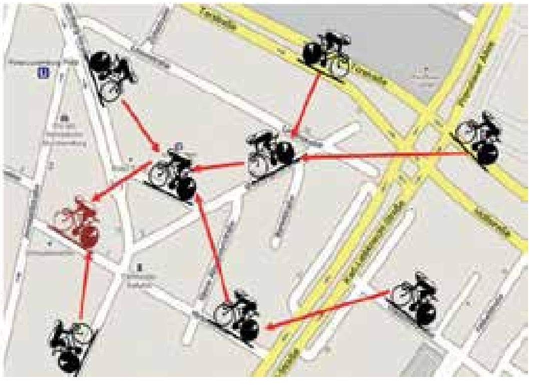


Figure 17. The BikeNet system architecture

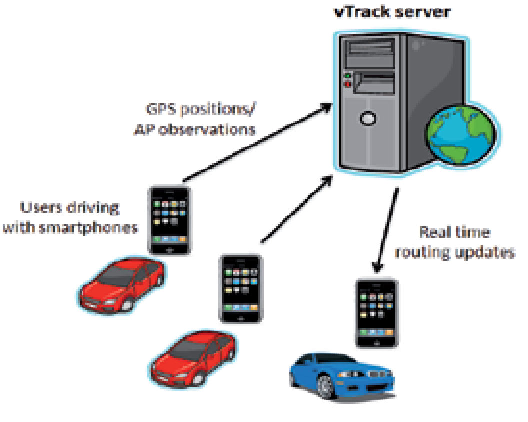


Figure 18. An illustration of traffic delay estimation application

5. Monitoring by using wireless sensors and actuators

Recent advances in pervasive computing, communication and sensing technologies are leading to the emergence of wireless sensor and actuator networks (WSANs). A wireless sensor and actuator network is a group of sensors and actuators that are geographically distributed and interconnected by wireless networks. Sensors gather information about the state of the physical world, and the actuators react to the information by performing appropriate actions. In most of the mission-critical monitoring applications, it also expects to respond to the sensed data and event by performing corresponding actions upon the system. WSANs enable the application systems to sense, interact, and change the physical world, e.g., to monitor and manipulate the temperature and lighting in a smart office or the speed and direction of a mobile robot. WSANs are widely used in mission-critical applications such as home automation, environmental monitoring and industrial control system, etc.

5.1. Architecture

5.1.1. Network-based architecture

There are two architectures called semi-automated and automated architectures for WSANs, respectively.

In semi-automated network architecture, one or more controller entities explicitly exist in the network. The controller is functional modules embedded in the base stations or separated nodes equipped with sufficient capabilities. The sensors collect the data and transmit to the controller. Then the controller produces control commands and sends them to the actuators according to the sensed information. The actuators perform the actions according to the commands they receive.

In automated network architecture, there is no explicit controller entity in the network. In this case, the collected data of sensors will be transmits directly to the actuators. The actuators will make decision and perform the actions according to the data sensed. In this kind of architecture, the actuators serve not only the executor of actions but also the controller.

5.1.2. Application-specific architecture

The architecture design problem is always an application-specific open issue. The concept of service-oriented sensor network architecture was proposed that encompass a set of sensor services and enables rapid scalable systems based on reusable services (Rezgui & Eltoweissy, 2007). For deploying this idea in WSANs need to handle problems such as the functionality interface and probabilities of service, as well as identifying the differences of sensor and actuator services.

There are usually QoS requirements for WSANs, the architecture needs to be designed with in mind the QoS demand and resource constraints of sensor and actuators respectively. The QoS-aware architecture design should consider the communication and coordination protocols within its framework, which is usually achieved by using cross-layer design to effectively optimize the network performance. The requirements imposed on WSANs are varied and depend on the applications. Considering mission-critical monitoring applications, WSANs should have high scalability to support timely data collection, event detection and task exectution. These requirments should be best addressed at the network architecture. The well investigated and appropirate architecture will benefit the network protocoals and other solutions.

5.1.3. Software-based architecture

In order for the wireless sensor and actuator networks to function, software-based architecture should be proposed and effectively manage unreliable dynamic distributed communication and coordination with the resource constraints, hostile environments and mission-critical tasks. The software-architecture can be achieved by middleware that can provide with a set of services and solutions involved in the network application (Chaczko et al., 2011). The software-based architecture is based on layered model, i.e., a set of services can interact with highly scalable central services. All the services such as node management service, task service, configuration service etc., are integrated through layers in the architecture.

There is another solution to design the software-based architecture based on collaborative operations between the sensors and actuators, enabling distributed sensing and action performing. The network functions are categorized such as: localization, communication, target tracking and standardization of services/interfaces. The data flow of sensors and actuators are categorized and analyzed. Then, the communication flow among network components: sensors and actuators are defined by the access method, physical link and formulation. The network software paradigm will be designed based on the function flow, data flow and communication flow.

5.2. Coordination

The coordination in WSANs is responsible for effective data sensing and action performing with limited resources by fully making use of node mobility. From the control perspective, mission-critical monitoring application systems are inherently real-time systems in the sense that control actions should be performed on the physical systems through valid node coordination. From the function perspective of coordination, it includes sensor-actuator coordination and actuator-actuator coordination, which is mainly responsible for network collaboration and task allocation. When a sensor gets the event packet, it will decide which actuator it reports to. Each actuator takes charge for its local real-time data collection from nearby sensors to form clusters in the monitoring filed, i.e., the actuator acts as a cluster head in the local cluster. And the clusters would vary with the dynamic topology, residual energy and reliability of mobile actuators. The objective of actuator-actuator communication and coordination is to select the best actuators to perform appropriate reactions towards the event. Each actuator acts as a collector by receiving event data from a subset of the sensors. However, one actuator acting as a data collector may not be able to act in its event area, i.e., the event area may not be totally covered by the action range of the actuator or the actuator has no enough energy to do so, etc.. For this reason, before action, actuator-actuator coordination is required to make decision on the choice of optimal actuators to perform tasks by considering the acting capability of actuators such as acting coverage, energy level and completion time.

5.2.1. Coordination in mission-critical environment

There is an optimal strategy for actuator coordination that is formulated as a mixed integer non-linear program. The objective of the optimization problem is to find minimal number and velocity of appropriate actuators to finish the task that resides in the surveillance area. Each occurring event e in the event space Ω can be characterized by $\{F(e), Pr(e), S(e), D(e)\}$,where $F(e)$ describes the event type, $Pr(e)$ the priority, $S(e)$ the event area, and $D(e)$ the action completion bound, i.e., the maximum allowed time from the instant when the event is collected to the instant when the associated action needs to be completed. The following notations are used in the problem formulation:

$X(e)$is a binary vector $[x_a^{(e)}]$ whose element is equal to 1, if actuator a acts on $S(e)$.
$V(e)$ is a vector $[v_a^{(e)}]$ whose element represents the velocity assigned to actuator a.
$T_a^{\Omega,(e)}$ is the time that actuator a needs to complete the task associated with event e when a is part of an acting team.
$d_a^{(w)}$ is the distance between actuator a and the center of the event area S when the acting range of actuator a is not within the event area.
$T_a^{M,(e)}$ is time need by actuator a to reach the event area.
T^c is the coordination time for actuators.
η_a^f is the performing rate (m^2/s) of actuator a acting on an event type f.
S_a^c is the subset of coordinating actuators when an event e occurs.

The problem can be cast as follows:

Find:

$$X^{(e)} = [x_a^{(e)}], V^{(e)} = [v_a^{(e)}] \tag{10}$$

Minimize:

$$\sum_{a \in S_a^c} x_a^{(e)} * v_a^{(e)} \tag{11}$$

Subject to:

$$T_a^{M,(e)} * v_a^{(e)} = d_a^{(w)}, \forall a \in S_a^c \tag{12}$$

$$0 \le v_a^{(e)} \le v_a^{\max} \tag{13}$$

$$T_a^{M,(e)} + T_a^{\Omega,(e)} + T^c \le D^{(e)} \tag{14}$$

$$\sum_{a \in S_a^{f,(e)}} x_a^{(e)} \geq 1 \tag{15}$$

$$\sum_{a \in S_a^c} x_a^{(e)} \cdot \eta_a^f \cdot T^{\Omega(e)} \geq S^{(e)} \tag{16}$$

The other promising ways to go are using feedback control, swarm intelligence and machine learning could be used to make design on coordination control algorithms. The distributed algorithms make design on heuristic methods in optimizing network performance with flexible resource management in dynamic and unpredictable environments.

5.2.2. Distributed task allocation

The distributed coordination in mission-critical monitoring, a negotiation-based algorithm can be utilized to make agreement among actuators about collaboration and task allocation.

When an actuator gets the last data packet of an event from local sensors, it broadcasts out a *Negotiate*(*actuator_id*, *Event_id*, *Event_priority*) message, and we call this actuator as negotiation invoker. The *Event_id* is the id of the event that the actuator collects; and the *Event_priority* is an application-specific identity used to discern the urgent degree of each event, such as "very high", "high", "medium", "low", and "very low". The event priority is predefined according to the application requirements. The node that receives a *Negotiate* message will reply an *ActReply*(*actuator_id*, *energy*, *location*, *Event_id*, *Event_priority*) message with its own id, event id and priority. The negotiation is executed in the local area of the actuator invoking the process. Multiple sensors placed in the monitoring field could be used to detect the same event, so one or multiple actuators will get the sensing data of the exact event. Through the process of backward *ActReply* message, the data of the same event (with the same *Event_id*) will be aggregated during the backward process and forwarded back to the first actuator that invokes the negotiation. If there is more than one event sensed by an actuator, then after the data aggregation for the same event, a multi-event ordered task assignment should be involved. The task assignment on this aggregation actuator will be executed according to the event priority, i.e., the event with higher *Event_priority* rank will win. If there is a tie of the event priority, the event with lower *Event_id* will win. Beyond the above, the aggregation actuator will get to know the whole area and location of the detected event through multi-actuator communication. The event area could be represented by using the left upside coordinate (Ev_X_i, Ev_Y_i) and the right downside coordinates (Ev_X_j, Ev_Y_j) approximately. During negotiation, we could combine each sub event area gathered by each actuator into the whole event area.

6. Summary

Considering the changing and time-varied topology and application scenarios in surveillance areas, the monitoring applications under mission-critical environment are very challenging. In this chapter, we have given some of the main concern for design in mission-critical monitoring application scenarios, and bring forward efficient methods and solutions according to network connectivity, dynamic application scenarios. Major research challenges and open research issues in mission-critical monitoring applications of WSNs are also outlined.

Author details

Yuanyuan Zeng and Deshi Li
School of Electronic Information, Wuhan University, Wuhan, China

Kai Xiang
School of Information Management, Hubei University of Economics
School of Computer, Wuhan University, Wuhan, China

Acknowledgement

This work was supported Chinese National Science Foundation Project (no. 61103218), Hubei Province National Science Foundation Project (no.2011CDB446), Wuhan University Automatic Research Project (no.111093), and Chinese National Program on Key Basic Research Project (no. 2011CB707106).

7. References

Poovendran, R. (2010). Cyber-physical systems: close encounters between two parallel worlds, *Proceedings of the IEEE*, 98(8), pp. 1363–1366.

Karp, B.; & T. Kung, H. (2000). GPSR: Greedy Perimeter Stateless Routing for Wireless Networks, *Proceedings of ACM MobiCom*, pp. 243-254.

Yu, Y.; Estrin, D.; Govindan, R. (2011) Geographical and energy-aware routing: a recursive data dissemination protocol for wireless sensor networks, UCLA Computer Science Department Technical Report, pp. 1-11.

Powell, O.; Nikoletseas,S. (2007). Geographic routing around obstacles in wireless sensor networks, Technical report, Computing Research Repository.

Fang, Q.; Gao, J.; Guibas, L. (2006).Locating and bypassing holes in sensor networks, *Mobile Networks and Applications*, 11(2), pp. 187-200.

Rogers, A.; David, E.; R. Jennings, N. (2010). Self-organized routing for wireless microsensor networks, *IEEE Transactions on Systems, Man, and Cybernetics*, 35(3), pp.349-359.

He, T.; Stankovic, J.; Lu, C.; Abdelzaher, T.(2003). SPEED: A Stateless Protocol for Real-time Communication in Sensor Networks,*Proceedings of ICDCS'03*.

Felemban, E.; G. Lee, C.; Ekici, E.; Boder, R.; Vural, S.(2005). Probabilistic QoS Guarantee in Reliability and Timeliness Domains in Wireless Sensor Networks, *Proceedings of InfoCom*.

Chipara, O.; He, Z.; Xing, G.; Chen, Q. (2006). Real-time Power-aware Routing in Sensor Networks, *Proceedings of 14th IEEE International Workshop on Quality of Service*.

Ahmed, A.; Fisal, N. (2008). A Real-time Routing Protocol with Load Distribution in Wireless Sensor Networks, *Computer Communications*, 31(14), pp.3190-3203.

Kim, J.; Ravindran, B. (2009). Opportunistic Real-time Routing in Multi-hop Wireless Sensor Networks, *Proceedings of the 2009 ACM Symposium on Applied Computing*.

Sivrikaya, F.; Geithner, T.; Thruong, C.; A. Khan, M.; Albayrak, S. (2009). Stochastic Routing in Wireless Sensor Networks, *Proceedings of IEEE International Conference on Communications Workshop*.

Yang, S.; Cheng, H.; Wang, F. (2010). Genetic Algorithms with Immigrants and Memory Schemes for Dynamic Shortest Path Routing Problems in Mobile Ad hoc Networks, *IEEE Transactions on Systems, Man, and Cybernetics*, 40(1), pp. 52- 63.

Li, Y.; Ai, C.; Vu, C.; Pan, Y. et al. (2010). Delay-bounded and energy-efficient composite event monitoring in heterogeneous wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 21(9), pp. 1373-1385.

Li, H.; Fen, L.(2009). Dynamic adaptive cooperative routing for emergency data in wireless sensor networks, *Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems*.

Byun, J. , Park, S. (2011). Development of a self-adapting intelligent system for building energy saving and context-aware smart services, *IEEE Transactions on Consumer Electronics*, 57(1), pp: 90-98.

Zeng, Y.; Sreenan, C.; Sitanayah, L.; Xiong, N.; Park, J.; Zheng, G. (2011). An emergency-adaptive routing scheme for wireless sensor networks for building fire hazard monitoring, *Sensors*, 11(3), pp.2899-2919.

Tseng, C. Y.; Pan, S. M.; Tsai, Y. Y. (2006). Wireless Sensor Networks for Emergency Navigation, *IEEE Computer*, 39(7), pp. 55-62.

Sobral M. & Becker, L. (2008). A Wireless Hybrid Contention/TDMA-based MAC for Real-time Mobile Applications, *Proceedings of SAC'08*.

Lu, G. & Krishnamachari, B.(2007). Minimum Latency Joint Scheduling and Routing in Wireless Sensor Networks, *Ad Hoc Networks*, 5(6), pp. 832-843.

Joseph, V.; Sharma, V.; Mukherji, U. (2009). Joint Power Control, Scheduling and Routing for Multihop Energy Harvesting Sensor Networks, *Proceedings of 4th ACM workshop on performance monitoring and measurement of heterogeneous wireless and wired networks*.

Jain, K.; Padhye, J.; Padmanabhan V. N; Qiu, L. (2003). Impact of interference on multi-hop wireless network performance, *Proceedings of ACM MobiCom'03*.

Hu J. & Wellman, M. P.(2003). Nash Q-learning for General-sum Stochastic Games, *Journal of Machine Learning Research*, 4, pp. 1039-1069.

Xing, G., Wang, X.; Zhang, Y.; Lu, C.; Pless, R.; Gill, C.(2005). Integrated coverage and connectivity configuration for energy conservation in sensor networks, *ACM Transactions on Sensor Networks*.

Gupchup, J.; Terzis, A.; Burns, R.C.; Szalay, A.S.(2009). Model-Based Event Detection in Wireless Sensor Networks, *Proceedings in Proceedings of CoRR*.

Zhuang, Y.; Chen, L.; Wang, X.; Lian, J. (2007). A Weighted Moving Average-Based Approach for Cleaning Sensor Data, *Proceedings in IEEE ICDCS*.

Rachlin, Y.; Negi, R.; Khosla, P. (2005). Sensing Capacity for Discrete Sensor Network Applications, *Proceedings in ACM/IEEE IPSN*.

Eriksson, J.; Girod, L.; Hull, B.; Newton, R.; Madded, S.; Balakrishnan, H.(2008). The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring, *Proceedings in ACM MobiSys*.

Benbasat A.& Paradiso, J. (2007).A Framework for the Automated Generation of Power-Efficient Classifiers for Embedded Sensor Nodes, *Proceedins in ACM SenSys*.

Kang, S.; Lee, J.; Jang, H.; Lee, H.; Lee, Y.; Park, S.; Park, T.; Song,J.(2008). SeeMon: Scalable and Energy-efficient Context Monitoring Framework for Sensor-right Mobile Environment, *Proceedings in ACM MobiSys*.

Singh, A.; Ramakrishnan, C.; Ramakrishnan, I.; Warren, D.(2008) A Methodology for In-Network Evaluation of Integrated Logical-Statistical Models, *Proceedings in ACM SenSys*.

Zappi, P.; Lombriser, C.; Steifmeier, T.; Farella, E.; Roggen, D.; Benini, L.;Troster,G. (2008).Activity Recognition from On-Body Sensors: Accuracy-Power Trade-Off by Dynamic Sensor Selection, *Proceedings in EWSN*.

Keally, M.; Zhou, G.; Xing, G.(2010). Watchdog: Confident Event Detection in Heterogeneous Sensor Networks, *Proceedings of IEEE RTAS*.

Keally, M.; Zhou, G; Xing, G.; Wu, J.; Pyles, A. (2011). PBN: Towards Practical Activity Recognition Using Smartphone –Based Body Sensor Networks, *Proceedings of ACM SenSys*.

Upadhyay, R.; Pringle, G.; Beckett, G.; Potter, S.; Han, L.; Welch, S.; Usmani, A.; Torero, J. L. (2009). An Architecture For An Integrated Fire Emergency Response System For The Built Environment, *Fire Safety Science*,9,pp. 427-438.

Eisenman S. B.et. al.(2007). The BikeNet Mobile Sensing System for Cyclist Experience Mapping, *Proceedings of Sensys'07*.

Thiagarajan A.; et. al. (2009). Vtrack: Accurate, Energy-Aware Road Traffic Delay Estimation using Mobile Phones, *Proceedings of Sensys'09*, pp.85-98.

Zeng, Y.; Murphy, S.; Sitanayah, L.; Tabirca, T.; Truong, T.; Brown, K.; Sreenan. C.J. (2009). Building Fire Emergency Detection and Response Using Wireless Sensor Networks, *Proceedings of 9th Information Technology & Telecommunications Conference (IT&T)*.

Rezgui, A. & Eltoweissy, M.(2007) Service-oriented sensor-actuator networks: Promises, challenges, and the road ahead, *Computer Communications*, 30(13), pp.2627-2648.

Chaczko, Z.; Chiu, C.; Aslanzadeh, S.; Dune, T. (2011). Software Infrastructure for Wireless Sensor and Actuator Networks, P*roceedings of 21st International Conference on Systems Engineering (ICSEng)*,pp. 474-479.

Chapter 12

End-To-End Monitoring System for the Preventing Deterioration in Art and Cultural Objects

Luca Bencini and Stefano Maddio

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48303

1. Introduction

A fundamental responsibility of a museum is the prevention of deterioration of art and artifacts through control of the environment in storage and exhibition. Preventive conservation entails storing, displaying, handling and maintaining a museum's collections in ways that promote long term stability and do not lead to deterioration. Thus, preventive conservation activities include monitoring the principal causes of deterioration, developing methods for secure display and storage, and ensuring the safety of works of art during their transport and loan to other museums.

Different types of collections need different forms of care. Many objects are composed of more than one material, each of which may respond differently to a variety of environmental factors. The major factors of deterioration are *light*, *temperature & humidity* and *pollutants*. Light is a form of energy that generates heat. Artifact deterioration is a result of chemical reactions that occur when an energy source changes the chemical structure of the object's surface. The amount of energy given off by a light source can be illustrated in the light spectrum. The electromagnetic spectrum is divided into wavelengths of energy, which range from low (radio waves) to high (gamma rays). The range of wavelengths from light sources (daylight and artificial light) can be divided into three regions: ultraviolet radiation or UV (300-400 *nm*), visible radiation (400-760 *nm*) and infrared radiation (over 760 *nm*). In essence, the shorter the wavelength of the light source, the more damaging to the surface of an object.

Temperature has a great effect on *Relative Humidity* (RH). Relative Humidity is a measure of the amount of moisture in the air relative to the amount the air is capable of holding, expressed as a percentage. If the air at a particular temperature contains half the water vapor it can hold at that temperature, the relative humidity is 50%. Acute changes in temperature and humidity will cause swelling and contraction as the materials in an object or artifact attempt to adjust to the environment. Objects are often composed of more than one type of material. Each material responds differently to water vapor in the air and adjusts to its particular *Equilibrium Moisture Content* (EMC) at different relative humidities. Of particular concern are the internal stresses created by expansion and contraction of the different materials as moisture diffuses into or out of the surrounding air.

Pollutants work in combination with other factors, such as temperature, relative humidity (RH) and light to cause deterioration. The museum environment poses a particular challenge because objects are often exhibited or stored in microenvironments, such as display cases or storage units. If the enclosure were made of a pollution emitting material, it would create a microenvironment in which the pollutants would remain confined with the objects.

Finally, the packing of art objects for shipment is another important aspect of preventive care for both stable and unstable objects. The extent of handling involved in packing the artwork, crating the artwork, movement during travel, and then unpacking the artwork could cause physical and chemical damage.

The most suitable technology to fit an invasive method of monitoring the environment without man attendance is a Wireless Sensor Network (WSN) system [1], [2], [5], [3], [4].

The requirements that adopting a WSN are expected to satisfy in effective cultural heritage monitoring concern both *system level* issues (i.e., unattended operation, maximum network life time, adaptability or even self-reconfigurability of functionalities and protocols) and *final user* needs (i.e., communication reliability and robustness, user friendly, versatile and powerful graphical user interfaces). The most relevant mainly concerns the supply of *stand-alone* operations. To this end, the system must be able to run unattended for a long period also in the absence of electricity. This calls for an optimal energy management ensuring that the energy spent is directly related to the amount of traffic handled and not to the overall working time. An additional requirement is *robust operative conditions*, which needs fault management since a node may fail for several reasons. Other important properties are *scalability* and *adaptability* of the network's topology, in terms of the number of nodes and their density in unexpected events with a higher degree of responsiveness and reconfigurability. Finally, several user-oriented attributes, including *fairness*, *latency*, *throughput* and enhanced data querying schemes [6] need to be taken into account even if they could be considered secondary with respect to our application purposes because the WSN's cost/performance trade-off.

Taking the before mentioned user and system requirements into account, a monitoring system based on WSN technology was developed. As Fig. 1 shows, it is comprised of:

- an *one level cluster tree WSN* endowed with sensing capabilities;
- an *Ethernet* or a *GPRS Gateway* (cluster head) for each cluster to gather data and provide a TCP-IP based connection toward a *Remote Server*;
- a *web application* which manages information and makes the final user capable of monitoring and interacting with the instrumented environment.

A cluster tree topology was adopted because it is the network architecture that best suits the structure of a museum. Moreover it allows to obtain simplicity, better performance in terms of latency, scalability and isolation of devices.

A GPRS Gateway was realized to extend the area of applicability of the WSN to zones in absence of wired connections such as outdoor environments.

More details of hardware, software and communication protocol design are provided in the next sections. In particular the chapter is organized as follows. Sections 2, 3 and 4 deal

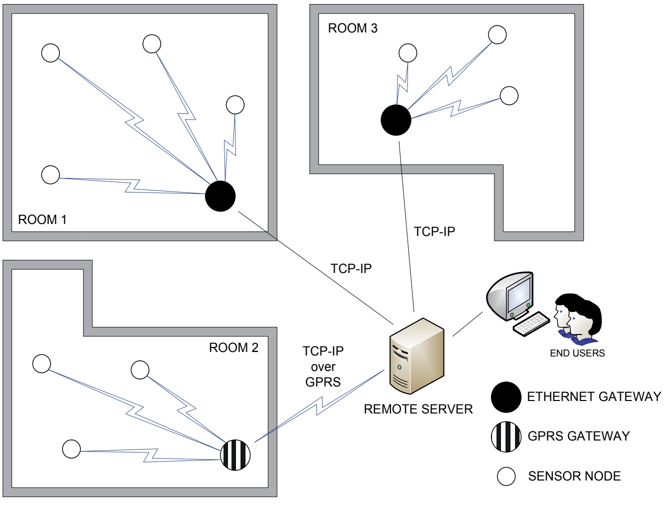


Figure 1. End-to-end WSN system

respectively with the system in terms of hardware, protocol and end user interface design. Section 2.3 is dedicated to a detailed description of the design of directional antennas. Section 5 describes the actual experiences, focusing on some case study analyses for highlighting the effectiveness and accurateness of the developed system. Finally, in Section 6 some conclusions are drawn in order to explain the future direction of the current research study.

2. Hardware and software design

Focusing on an end-to-end WSN system architecture, every constitutive element has to be selected according to the application requirements and scenario issues, especially the hardware platform. Many details have to be considered, involving the energetic consumption of the sensor readings, the power-on and power-save states management and a good trade-off between the maximum radio coverage and the transmitted power.

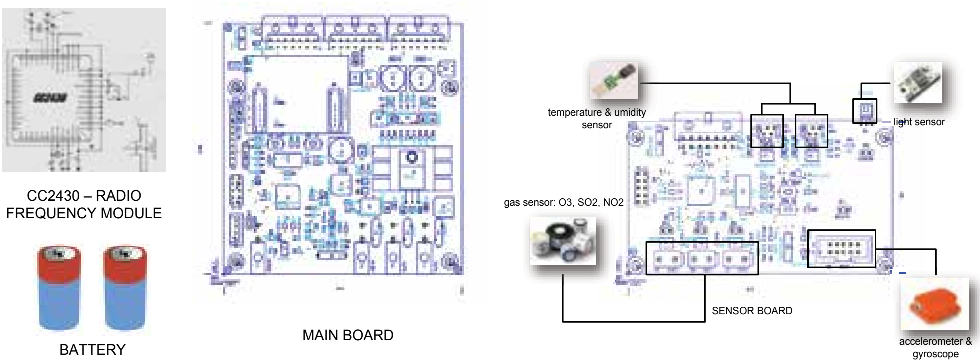
In the next subsections a detailed hardware description of the WSN components is provided.

2.1. Sensor node design

In order to allow greater flexibility in the placement of the network devices, each sensor node was realized separating the sensor unit (*Sensor Board*) from the power board and the communication unit (*Main Board*).

The Sensor Board is responsible for data acquisitions. As Fig. 2(a) shows, it can manage simultaneously two digital temperature-moisture sensors, one analogical light sensor, three analogical gas sensors (O_3, NO_2, SO_2) and one accelerometer-gyroscope sensor that is hardwired to the Sensor Board by a RS232 serial port. The Sensor Board recognizes automatically the sensors once they are plugged and sends Transducer Electronic Datasheets (TEDS) through the network up to the server, making an automatic sensor recognition possible by the system.

Table 1 provides a list of the sensor typologies used in our tests and applications.



(a) Sensor Board (b) Main Board

Figure 2. Sensor Node

MANUFACTURER	MONITORED PARAMETER
SENSIRION - SHT75 [7]	Humidity & Temperature
MicroSemi - TAOS TSL252R-LF [8]	Light
City Tech - EZT3NDH [9]	NO_2
City Tech - EZT3SH [9]	SO_2
City Tech - O33E1F [9]	O_3
XSens - Mti [10]	Accelerometer & Gyroscope

Table 1. Sensor Typologies

The Main Board handles the communication with the Sensor Board and with the gateway and manages the charge and discharge of the batteries. As Fig. 2(b) shows, it is composed by a RF module equipped by a omnidirectional antenna, a master board and a power board. The power board includes ten 3200 mAh @ 1.8 V rechargeable batteries connected in series.

Some pieces of information about the power consumption and the lifetime of the sensor nodes are provided in Section 3 where the communication protocol is described.

The Main Board can support and manage up to a maximum of three Sensor Boards thanks to a RS422 serial bus.

The "core" of each sensor node resides in the software/firmware modules developed within the chips of the master board (*ATMEL ATmega644P microcontroller*), RF board (*Texas Instruments CC2430 System-On-Chip*) and Sensor Board (*ATMEL ATmega644P microcontroller*). In fact these C/assembler modules permits to manage high and low power states and the charge and discharge of the batteries, to realize finite state machines, to query sensors at fixed intervals and to achieve anti-blocking procedures in the case of software failures or deadlocks. An important role is played by the communication module that allows the information exchange between each node and the gateway.

2.2. Ethernet and GPRS Gateway

The Ethernet Gateway is a Main Board hardwired with an *AK-NORD XT-Nano ethernet module* [11] through a RS232 serial interface.

It is equipped by a 2.4 *GHz* switched beam antenna (see Section 2.3) and by ten 3200 *mAh* @ 1.8 *V* rechargeable batteries connected in series that provide emergency power for 212 hours when the input power source, the utility mains or the *Power over Ethernet* (PoE), fails.

Data between the Ethernet Gateway and Remote Server are carried out over TCP-IP communication and encapsulated in a custom protocol; from both local and remote interfaces it is also possible to access part of the Gateway's configuration settings. To avoid data lost caused by link failures, a 128 *KB* SRAM memory is also mounted on the board to allows for data buffering.

The GPRS embedded Gateway is a stand-alone communication platform designed to provide transparent, bi-directional wireless TCP-IP connectivity for remote monitoring. In conjunction with *Remote Data Acquisition* (RDA) equipment, such as WSN, it acts when connected with a sensor node or when directly connected to sensors and transducers.

The main hardware components that characterize the gateway are:

- a 2.4 *GHz* switched beam antenna (see Section 2.3);
- a miniaturized GSM/GPRS modem, with embedded TCP/IP stack [14] [16];
- a powerful 50 *MHz* clock microcontroller responsible for coordinating the bidirectional data exchange between the modem and the master node to handle communication with the Remote Server;
- an additional 128 *KB* SRAM memory added in order to allow for data buffering, even if the wide area link is lost;
- several A/D channels available for connecting additional analog sensors and a battery voltage monitor.

Since there is usually no access to a power supply infrastructure, the hardware design has also been oriented to implement low power operating modalities, using a 12 *Ah* @ 12 *V* rechargeable lead battery and in addition a 20 *W* solar panel when the gateway is deployed in outdoor environment.

Data between the GPRS Gateway and Protocol Handler are carried out over TCP-IP communication and encapsulated in a custom protocol; from both local and remote interfaces it is also possible to access part of the Gateway's configuration settings. The low-level firmware implementation of communication protocol also focuses on facing wide area link failures. Since the gateway is always connected with the Remote Server, preliminary connectivity experiments demonstrated a number of possible inconveniences, most of them involving the *Service Provider Access Point Name* (APN) and *Gateway GPRS Support Node* (GGSN) subsystems. In order to deal with these drawbacks, custom procedures called *Dynamic Session Re-negotiation* (DSR) and *Forced Session Renegotiation* (FSR), were implemented both on the gateway and on the CMS server. This led to a significant improvement in terms of disconnection periods and packet loss rates.

The DSR procedure consists in a periodical bi-directional control packet exchange, aimed at verifying the status of uplink and downlink channels on both sides (gateway and CMS). This approach makes facing potential deadlocks possible if there is asymmetric socket failure, which is when one device (acting as client or server) can correctly deliver data packets on the

TCP/IP connection but is unable to receive any. Once this event occurs (it has been observed during long GPRS client connections, and is probably due to Service Provider Access Point failures), the DSR procedure makes the client unit to restart the TCP socket connection with the CMS.

Instead, the FSR procedure is operated on the server side when no data or service packets are received from a gateway unit and a fixed timeout elapses: in this case, the CMS closes the TCP socket with that unit and waits for a new reconnection. On the other side, the gateway unit should catch the close event exception and start a recovery procedure, after which a new connection is re-established. If the close event should not be signaled to the gateway (for example, the FSR procedure is started during an asymmetric socket failure), the gateway would anyway enter the DSR recovery procedure.

In any case, once the link is lost, the gateway unit tries to reconnect with the CMS until a connection is re-established.

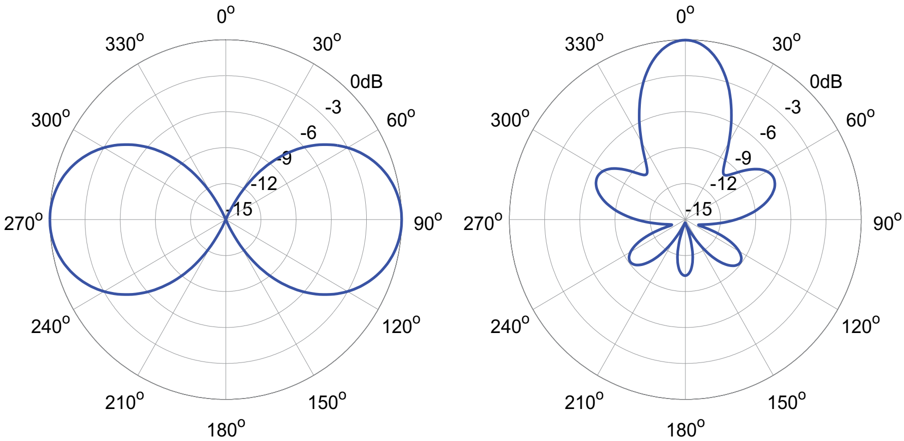
2.3. Antenna design

The typical basic antenna equipped in the sensor nodes of a WSN are *omnidirectional*, either in the form of printed IFA or as compact stylus. In this context, the term omnidirectional refers to a 2D isotropic behavior, as the canonical dipole antenna, radiating almost uniformly only in a transverse plane and affected by strong transmission zeros above and below. This kind of antenna is suitable for a isotropic broadcast communication (Figure 3(a)).

A *directional* antenna is a radiator specifically designed to be *directive* (Figure 3(b)). Directive antennas privilege the electromagnetic link toward specific direction, showing an high transmission gain peak. Directive antennas can improve the performance of WSNs in several ways. With a directional antenna the same received power is obtained with less transmitted power, achieving a better power efficiency - unicast communication. Alternatively, greater transmission range are possible with the same available power. With limited angular coverture, it operates as a spatial filter, which is an aid to contrast the interferences in a electrically polluted are, with significative advantages in term of link quality. Finally, the spatial re-usability can increase communication capacity and throughput. Nevertheless, a single antenna cannot be suitable for both broadcast and unicast links.

A *Switched Beam Antenna* (SBA) is a smart antenna able to establish a set of predetermined directional beams. An SBA is composed of a set of elementary radiators combined with a selection logic driven by a digital control activated by the node intelligence. While limited by the available sectors, switched beam antenna are simpler and cheaper than adaptive beamformers - antenna array able to generate beam in arbitrary direction - while maintaining similar advantages. The simplest SBA is made by a Single Pole N Through (SPNT) RF switch directly connected to an array of N printed antennas, whose geometry is arranged in order to cover the entire domain of interest.

Printed (patch) antennas are the most suitable radiators for SBA. Compatible to the standard PCB photo-etching technology and materials, at the price of a little performance drop, patches can assume any arbitrary planar shape, making them very versatile in terms of nominal resonant frequency, bandwidth, polarization and pattern shaping. When operating in fundamental mode, they typically exhibit a mono-lobe radiation pattern characterized by

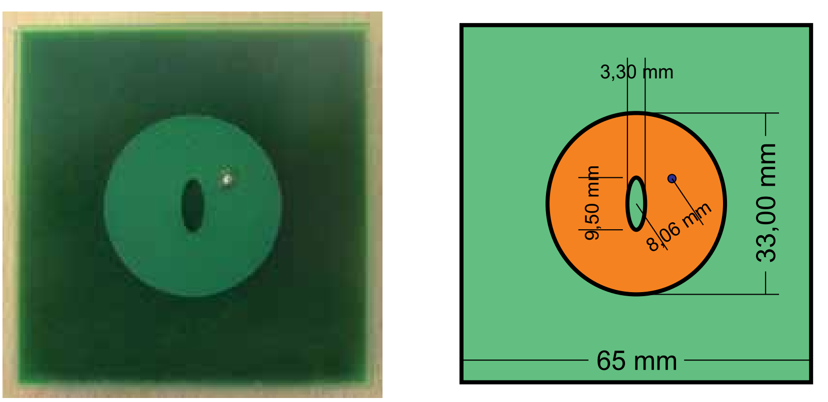


(a) Omnidirectional Pattern - Broadcast (b) Directional Pattern - Unicast

Figure 3. Typical Antenna Patterns

a radiation peak in the normal direction of the antenna plane (broadside radiator) with a maximum directivity in the range of 4-7 dB.

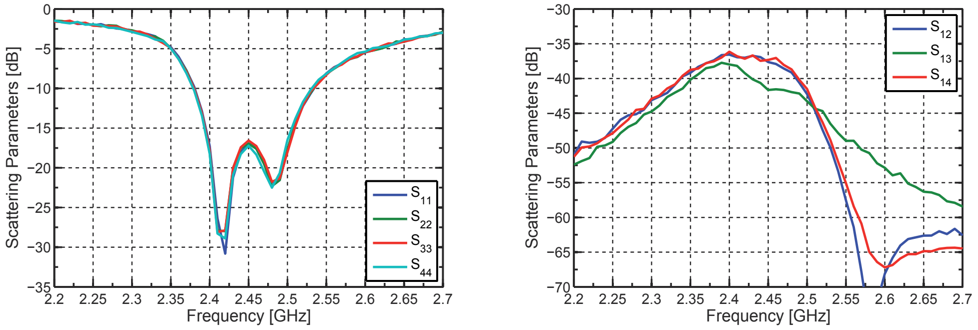
The SBA employed in the described WSN is a *Four Switched Beam Antenna* (FSBA), a device composed of four coaxially fed antennas arranged in a cubic structure as shown in Figure 6. The resulting elementary antenna layout designed in common FR4 substrate ($\epsilon_r = 4.4$, $h = 1.6$ mm, $\sigma_{Cu} = 5.8\, mS$) and its dimensions, along with the prototype photograph, are shown in Figure 4(b) Each antenna element ground is shaped in a $65\, mm \times 65\, mm$ square. The radiative system, while bulky, can operates as a mechanical shelter of the senor node.



(a) Element Photograph (b) Element Layout With Quotes

Figure 4. Design and Photograph of the Elementary Antenna of the SBA

The antenna element of the proposed system are compact directive patch antennas operating in Circular Polarization (CP), which ensure communication regardless of relative orientation. CP is also as an effective aid to combat the multipath effect, since the radiated field by a LHCP antenna inverts its rotation sense after reflecting on the ground, becoming a RHCP field, thus virtually invisible to a co-polarized LHCP receiving antenna. The Antenna design is based on the modal degeneration phenomenon, a smart way to achieve and control the CP effect without the aid of an external splitter. Following this approach, compact directive antenna design is possible [12, 13].



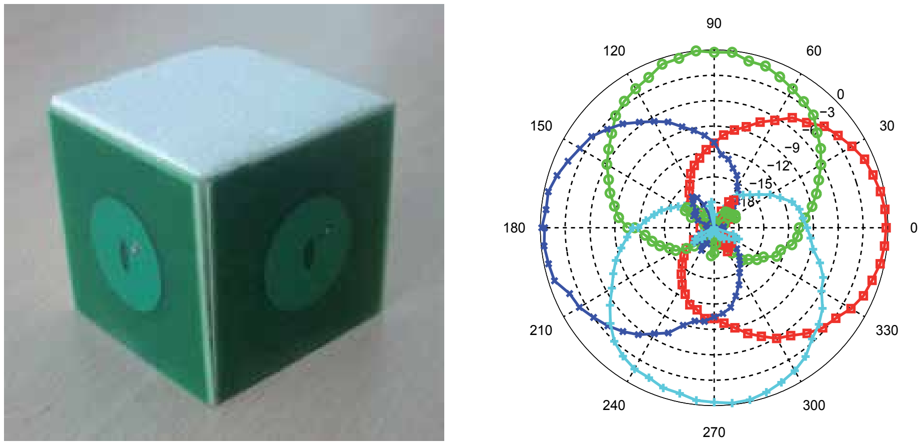
(a) Reflection Coefficient (S_{xx}) (b) Isolation (S_{xy})

Figure 5. Scattering Parameters of the Antenna Elements

Figure 5 shows the performance of the four antennas. The reflection coefficients are plotted in Figure 5(a), where a minimum around the center of ISM frequency confirms a good matching, quantified in a Return Loss (RL) of 17 dB and a 10 dB RL bandwidth exceeding 150 MHz. All the four traces exhibit the same dual min-max-min behaviour around the central frequency, a hint of the archived circular polarization in the area between the minimums. Dimensions and performance were traded off so as to be compact but at the same time without sacrificing performance. Figure 5(b) shows the isolation of the four antennas, confirming a minimum value of 36 dB, suitable for the application in exam.

The assembled structure composed of the four antennas shows a set of 4 individual directive pattern with a maximum over the four cardinal direction, individuated by the side of the box structure, and an Half Power Beam Angle of almost 90 degrees. The latter characteristic grants that the *cumulative pattern*, the pattern composed by the maximum of the four antennas, is almost isotropic, ensuring the communication coverture the entire 2π radian domain. The measurements azimuthal pattern at the design frequency of 2.45 GHz reveals a good Left Hand with a Co/Cross discrimination between 18 and 21 dB depending on the antenna. Optimal CP is achieved in the broad-side direction, while the back lobe experiments a CP inversion, a typical behavior of the modal degenerated antenna, which is an aid for the link discrimination. The absolute gain value is estimated as 3.85 dB, with an estimated antenna efficiency of, compatible with the sub-optimal performance of the cheap substrate.

In Figure 6 is depicted a Four Beam Antenna (FBA), proposed as the basic node of the WSN. The four sector beams of the 4BA are also shown. The radiation patterns are measured in operative condition, with the antenna elements connected to the switch. The maximum gain is estimated as 2 dB, a low value affected by the switch insertion loss and unavoidable co-channel coupling. The advantages in generating 4 beams is that the cumulative pattern can cover the entire 2D angular range with a cubic box arrangement, but dedicated links can be established toward the four directions.



(a) Antenna Arrangement (b) Antenna Beams

Figure 6. Four Switched Beam Antenna Photograph and Radiative Performance

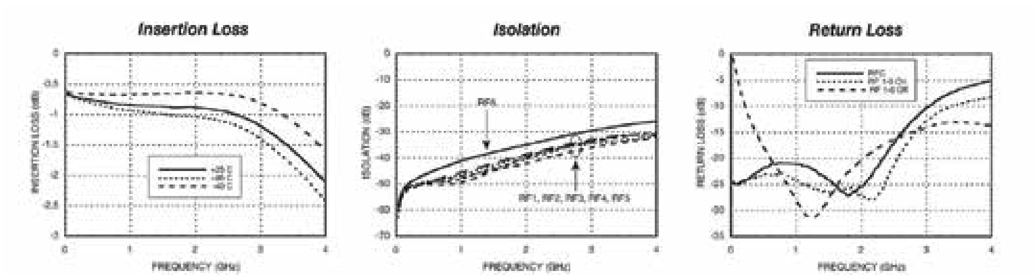


Figure 7. SP4T Characteristics Taken from Hittite Datasheet

The selection mechanism of proposed SBA is a non-reflective SP4T switch by Hittite, model HMC252QS24. The switch performance, reported in Figure 7, are suitable for the application in exam. The adoption of this kind of switch minimizes the coupling between the 4 radio channels ensuring a minimal pattern and polarization corruption.

3. Communication protocol design

Apart the Ethernet Gateway, all devices are energy constrained. Therefore smart power saving procedures have to be adopted to increase network lifetime. Some common approaches are the management of both sleep and active states, the on board integration of directional antennas and their integration within the communications framework. As these aspects belong to both the Physical (PHY) and Medium Access Control (MAC) layers, they might be integrated to achieve an overall energy efficiency, managing, for example, at the same time the duty cycle and the beams of the on board directional antennas. The way to accomplish this goal effectively relies on the so called *cross-layer* protocol design principle [15].

According to this principle, the communication protocol was designed and realized.

In the following sections the protocol stack is described and its performance are provided.

3.1. Physical and MAC layer

The Physical (PHY) layer and in part the Medium Access Control (MAC) layer are hardware implemented by the *Texas Instruments CC2430 System-On-Chip* in accordance with IEEE 802.15.4 standard [17].

The Physical layer (PHY) provides the data transmission service and performs channel selection and energy and signal management functions. It operates on 2400 - 2483.5 MHz unlicensed frequency bands with a rate of 250 $Kbps$. Moreover, it employs a 16 - ary quasi-orthogonal modulation technique, in which each four information bits are mapped into a 32 - chip pseudo-random noise (PN) sequence. The PN sequences for successive data symbols are then concatenated and modulated onto the carrier using an Offset Quadrature Phase Shift Keying (O-QPSK).

The Medium Access Control (MAC) layer implemented by the *Texas Instruments CC2430 System-On-Chip* allows the transmission of MAC frames through the use of the physical channel. Besides the data service, it also controls frame validation, guarantees time slots and offers hook points for secure services. Finally it implements CSMA-CA mechanism for channel access.

To improve the energy efficient of the system, a MAC protocol based on *sleep* and *active* states [18] [19] and IEEE 802.15.4 features and able to manage switched beam antennas was developed.

According to it, each device wakes up independently, entering an initial idle state (*init state*) in which it remains for the time interval necessary for performing the elementary CPU operations and to be completely switched on (T_{init}). Moreover, before entering the *discovery state*, each device starts to organize the time into frames whose durations areT_f .

In the *discovery state* each sensor node tries to associate itself with a gateway and to establish a time synchronization with it. Vice versa each gateway tries to build up its cluster of sensor nodes.

Each gateway remains in a listening mode for a time interval equal to $T_{set-up} \geq 2T_f$ and begins to periodically broadcast a HELLO message to each angular sector (i.e., the coverage area within a certain side lobe) sending its *ID* and its *phase*. The *phase* is the time interval after which the sender exits from the *discovery state*, enters the *regime state* and changes back in listening mode in that particular sector.

A sensor node that receives a HELLO message from a gateway adds it to the list of its own active gateways and transmits an acknowledgement. At the end of the *discovery state*, each sensor node chooses only one gateway as cluster head. Even so it keeps trace of the others for backup.

A gateway that receives a HELLO message from another gateway discards instead the information.

Each sensor node performs the same procedure. Since it is equipped by an omnidirectional antenna (dipole antenna), it has not to worry about changing the antenna's angular sector.

A sensor node that receives a HELLO message from another sensor node discards the information. A gateway that receives a HELLO message from a sensor node adds it to the list of its own active sensor nodes.

Once the *discovery state* has expired, each device enters the *regime state*. Within this state the operation mode is duty cycled with a periodic alternation of listening and sleeping sub-periods whose time intervals are T_l and T_s respectively. The duty cycle function is given by the following formula:

$$d = \frac{T_l}{T_l + T_s} \tag{1}$$

In the *regime state* each device tries to preserve the synchronization established during the *discovery state*. To this purpose each gateway sends a frame-by-frame HELLO message in a unicast way to the sensor nodes in its list belonged to different sectors according to the phase transmitted by them in previous HELLO messages. The same procedure is performed by each sensor node toward the gateway each one is associated with. Fig. 8 shows the synchronization messages exchange.

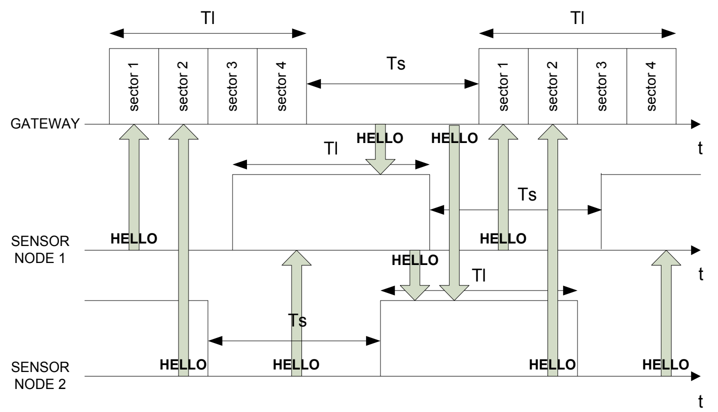


Figure 8. MAC Protocol Synchronization Messages Exchange

As in the *discovery state*, the HELLO message contains the *ID* and the *phase* that, in this case, is the time interval after which the sender claims to be again in the listening status waiting for the HELLO messages. The *phase* ϕ is evaluated according to the following rule:

$$\phi_1 = \tau - T_l \tag{2}$$

if the node is in the sleeping mode, where τ is the time remaining to the beginning of the next frame. Conversely, if the node is in the listening status, ϕ is computed as:

$$\phi_2 = \tau + T_s \tag{3}$$

To complete the protocol characterization, a device turns off entering the *off state* when its battery is depleted.

If a gateway presents an improper functioning or its battery is depleted, the sensor nodes belonging to its cluster are not capable of sending data to the Remote Server. To this purpose, a *recovery state* was introduced.

Each sensor node monitors every frame the *link quality* (LQ) defined as:

$$LQ = \frac{N_{Rx-GW}}{N_{Tx-SN}} \quad (4)$$

where N_{Tx-SN} represents the number of HELLO messages sent by a sensor node and N_{Rx-ON} is the number of ACKs sent by the associated gateway. When the value of LQ is below a certain threshold (LQ_{th}), the sensor node wakes up and starts to broadcasts periodically SOS messages until a gateway answers or a recovery time set a priori is elapsed. If a gateway receives this message it includes the orphan sensor node in its cluster sending a HELP message. At the end of the procedure the respective tables are updated.

If there are no gateways able to support an orphan sensor node (i.e., HELP messages is not received), the network tries to establish an end-to-end path between the sensor node and a remote gateway. In particular, the orphan sensor node broadcasts periodically SOS messages to the sensor nodes in its coverage area. The recursive application of this flooding procedure sets up a sort of ad-hoc network. The orphan sensor node will choose the path with the minimum number of hops.

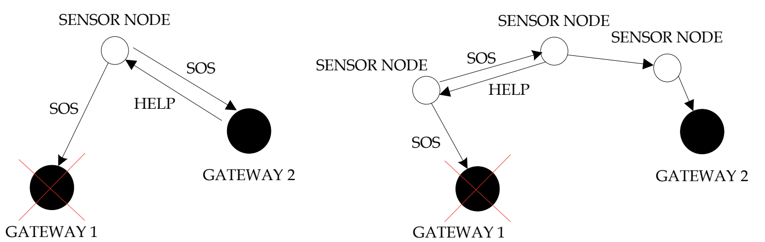


Figure 9. Recovery State

Both the procedures are shown in Fig. 9.

3.2. Application layer

Since higher-level layers and interoperability sub-layers are not defined in the IEEE 802.15.4 standard, a custom protocol was developed to attain a full interaction between the final user and the WSN. The application protocol is composed by:

- CONTROL MESSAGES (i.e., ping, ack), to verify the network status;
- MANAGEMENT MESSAGES, to access part of network element configuration settings and to change some important monitoring parameters;
- DATA MESSAGES, to receive row data from the WSN;
- ERROR MESSAGES, to detect system failures due for example by low battery levels or communication problems.

3.3. Performance analysis

In order to evaluate the performance of the proposed communication protocol in terms of power consumption and latency, some simulations were performed. The simulated system was developed by means of a network protocol simulator called *NePSing*, that is, a C++ framework specifically designed for modeling the evolution of time-discrete asynchronous networks [20]. Finally, to validate the simulations, the results were compared with those obtained by real measurements.

The most relevant simulation parameters are summarized in Table 2.

Sensor Node	
energy consumption (active mode)	60 mA
energy consumption (sleep mode)	0.7 μA
sleeping sub-period $[T_s]$	5 s
Ethernet Gateway	
energy consumption (active mode)	60 mA
energy consumption (sleep mode)	1 μA
sleeping sub-period $[T_s]$	5 s
GPRS Gateway	
energy consumption (active mode)	200 mA
energy consumption (sleep mode)	1 mA
sleeping sub-period $[T_s]$	5 s

Table 2. Simulation Parameters

The adopted network structure is shown in Fig. 10. It is comprised of a GPRS Gateway, an Ethernet Gateway and twelve sensor nodes deployed in a 50 m x 50 m area.

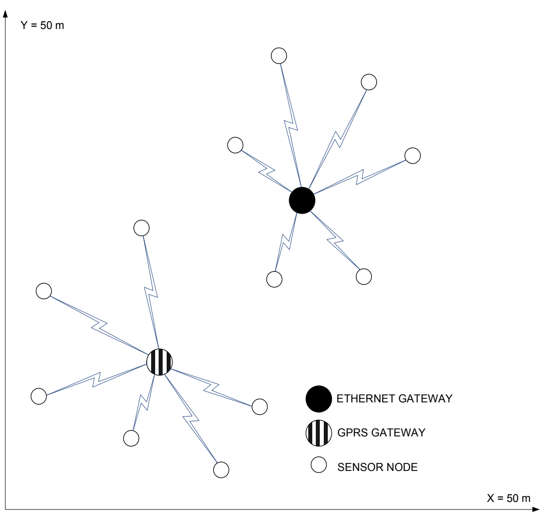
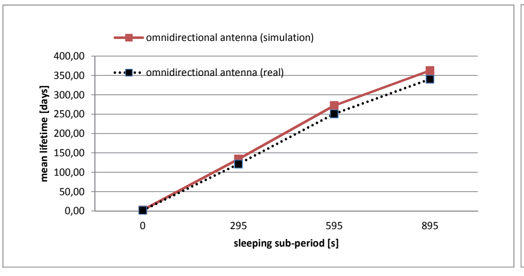


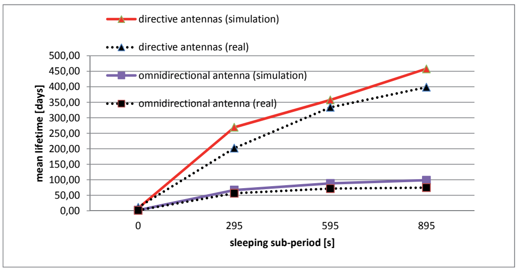
Figure 10. Adopted Network Topology

The adopted antenna model is an ideal smart antenna formed by a group of four non overlapping adjacent beams that cover the omnidirectional area. In particular, an antenna radiates in a fixed sector of $\pi/2$ radians, thus providing an increased gain over a restricted range of azimuths respect to an omnidirectional antenna.

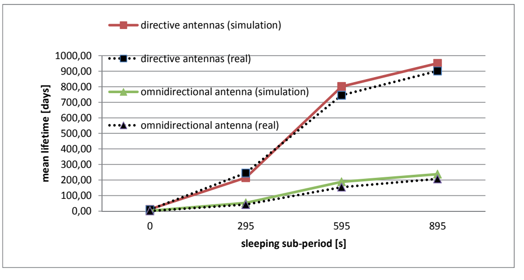
To give an insight on the protocol energy efficiency in Fig. 11 the mean lifetime of the network nodes (sensor nodes, Ethernet Gateway and GPRS Gateway) as a function of sleeping sub-periods is shown. When the nodes are *always on* ($T_s = 0$), the network lifetime corresponding to the sensor node lifetime is about two days. This time is sufficient to monitor continuously (every 15 s) an art object that is transfered from a museum to another. The introduction of a duty cycle ($T_s \neq 0$) and the use of directive antennas installed on gateways reduce the mean power consumption and increase the mean battery time. For example, the lifetime increment due to the installation of directive antennas on the GPRS Gateway is about 713 days in the case of $T_f = 15\ min$ ($T_s = 895\ s$). Finally, figures highlight the good accurancy of the simulation model by comparing mathematical predictions with real results.



(a) Sensor Node Lifetime



(b) Ethernet Gateway Lifetime



(c) GPRS Gateway Lifetime

Figure 11. Mean Lifetime vs. Sleeping Sub-Period

To conclude the analysis, the mean latency is taken into the account. The mean latency is the time interval between an orphanage detection due to a gateway fault and the consequent sensor node's association with a new gateway. In Fig. 12 the mean latency as a function of the number of sensor nodes is shown. The network nodes are *always on*. The link quality threshold (LQ_{th}) is set to 2. Firstly, it is important to note that the latency is very low thus underlying a low collision probability. It is about 10 s when the deployed sensor nodes are 6. Secondly, the mean latency is a linear function of the number of deployed sensor nodes, thus highlighting a good network scalability.

4. End user interface design

The *Remote Server* stores, processes and presents the information gathered by the WSN. Data are comprised of sensing (measures, battery level) and control/management messages. The final user may check the system status through graphical user interface (GUI) accessible via web. After the log-in phase, the user can select the proper monitored scenario (i.e., a museum,

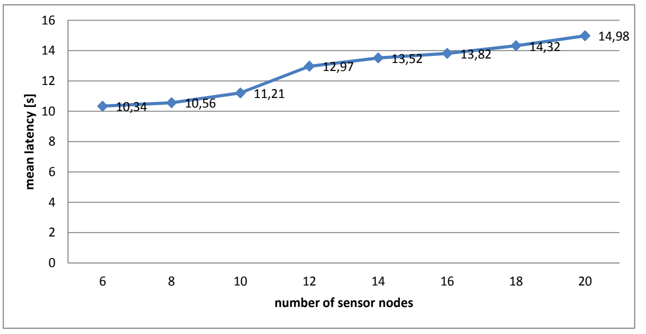


Figure 12. Mean Latency vs. Number of Sensor Nodes

a particular room or a particular art object). For each scenario the deployed WSN together with the gateway or part of WSN is schematically represented through an interactive map. In addition to this, the related sensors display individual or aggregate time diagrams for each node with an adjustable time interval (Start/Stop) for the observation. System monitoring could be performed both at a high level with a user friendly GUI and at a low level by means of message logging.

Afterwards two GUI examples are briefly described.

In Fig. 13 the frequency distribution and the cumulative frequency of the relative humidity are represented. The *frequency distibution* is the number of occurrences of a repeating value of a phenomenon per observation time. The *cumulative frequency* is the frequency of occurrence of values of the phenomenon less than or equal a reference value. For example the figure shows that during the observation period 1410 monitored samples of the relative humidity have assumed values within the interval between 40% - 45%. While 78% of monitored values are less or equal than 45%.

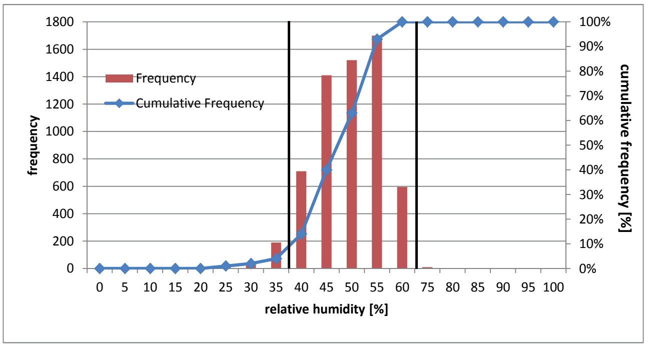


Figure 13. Frequency Distribution and Cumulative Frequency of the Relative Humidity

In Fig. 14 a temperature and relative humidity matrix obtained in 10 observation days and the related global *Performance Index* (PI) are shown. The PI is defined as the percentage of time in which the measured parameter lies within the required (tolerance) range. In this case the tolerance ranges are [40% - 60%] and [19C° - 24C°] for the relative humidity and the temperature respectively. The evaluated PI is 96.3%.

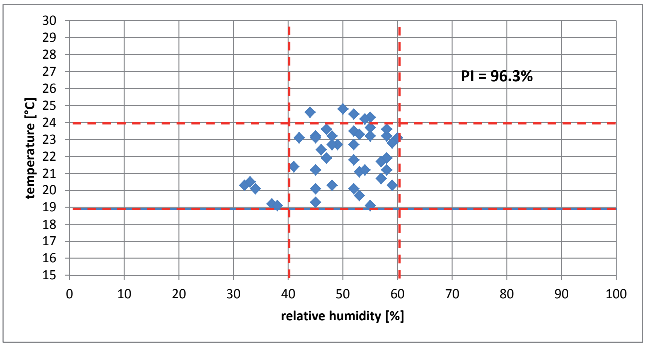


Figure 14. Global Performance Index on Temperature and Relative Humidity Matrix

5. Real experiences

The WSN system described above was tested and deployed in our laboratory and in one pilot site. The collected data represents a database of information on changes of the environmental parameters considered the principal causes of deterioration of art and artifacts.

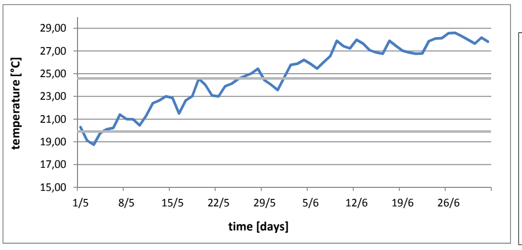
5.1. System tests

On May 1st 2009 in two rooms of our laboratory the first deployment was performed in order to evaluate the protocol performance and the capability of the system to provide useful and essential informations on changes of the principal environmental parameters. In fact the transition from Spring to Summer represents the most critical phase of the year from the point of view of climatic oscillations.

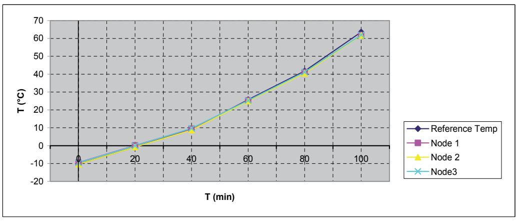
Eight sensor nodes equipped with temperature-humidity and light sensors and one Ethernet Gateway were installed in strategic points of the rooms.

To give an insight on the protocol efficiency, the Message Delivery Rate (MDR) was evaluated as the ratio between the messages correctly received by the remote server and the expected transmitted messages. After five months each sensor node showed a MDR over 97% in the case of a sampling interval equal to 10 minutes and over 94% in the case of 2 minutes. This confirms the robustness of the network and the reliability of the adopted communications solution. Moreover, the plotting data on the test period gave evidence of the usability and the utility of the proposed system. Fig. 15(a) shows that the environmental temperature of the rooms was too high for the optimal maintenance of art and artifacts. In particular from the end of May the temperature rose till the achievement of $28,6C^{\circ}$ thus exceeding the tolerance threshold. Conversely, the relative humidity values remained for most of the time within the optimal conservation range ($40 \leq RH \leq 60$) as Fig. 13 highlights.

In order to evaluate the accuracy of the temperature sensors (SENSIRION - SHT75), another important test was performed in the climatic chamber of the Microelectronic Laboratory of the University of Florence. To this end, a simple network composed by three sensor nodes and one GPRS Gateway was set up within the chamber. As the Fig. 15(b) shows a mean square error equal to 0.1 was evaluated from the comparison between the values detected by the sensor nodes and the reference temperature. This error is low respect to that obtained during the calibration of the commercial devices.



(a) Temperature Time Diagram

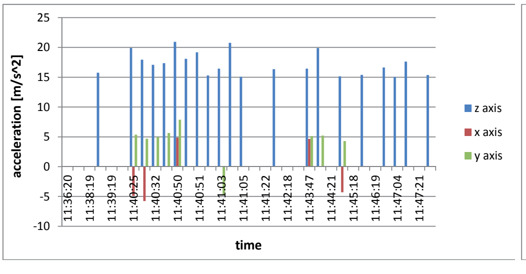


(b) Climatic Test Chamber

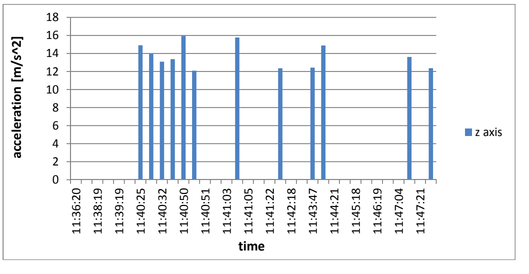
Figure 15. Temperature Tests

In order to evaluate the capability of the system to provide real time informations on the accelerations of moving objects, a road test was performed. Two sensor nodes equipped with one accelerometer-gyroscope sensor were installed inside and outside a container respectively. The container was placed within a van together with a GPRS Gateway.

Fig. 16(a) and Fig. 16(b) show the acceleration values over a fixed threshold, 4.905 m/s^2 for the x and y components and 11.829 m/s^2 for the z component. It is important to note that in the time interval between 11:36 a.m. and 11:47 a.m the container was subjected to considerable ripples caused by a dirt road. The container instead damped the oscillations.



(a) Outside the Container



(b) Inside the Container

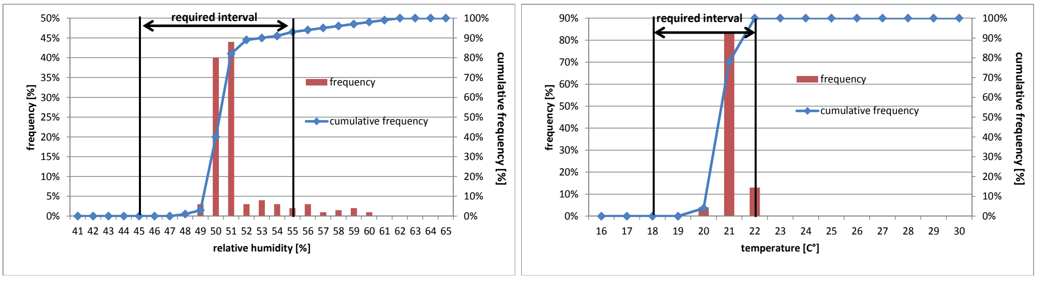
Figure 16. Acceleration Values

5.2. Pilot site description

The WSN system was deployed to qualify the thermal quality and the presence of gases during a temporary exhibit (paintings) in an Italian museum, throughout the entire heating season (from October to April). The analysis focused on the observation of thermal and hygrometric and gas data. The museum was equipped with two WSNs, controlling both air temperature, relative humidity and carbon dioxide inside and outside the showcases. The exhibit took place on one whole floor of the museum building, across two different exhibit areas. The exhibit floor area was about 1800 m^2 and the average room height was 3.5 m. The rooms were conditioned by an all-air system, working continuously 24 h a day.

The monitoring campaign was performed using 38 sensor nodes (including the gateways) with 107 sensors. These sensor nodes were located in each room, in order to measure and send values every 10 min.

Analysing the thermal quality, a prime importance was given to frequency distribution and cumulated frequency evaluation. Statistical values, frequency distribution, and cumulated frequency are shown, respectively, in Fig. 17(a) and Fig. 17(b). 94% of the relative humidity values are within the required interval of 45-55% (which leads to a PI of 94%), and most of the data are between 49% and 52%. 100% of the temperature values are within the required interval (which leads to a PI of 100%), and most of the data are between $21C^{\circ}$ and $22C^{\circ}$.



(a) Relative Humidity (b) Temperature

Figure 17. Frequency Distribution and Cumulated Frequency

A new global PI was introduced and calculated, considering both temperature and relative humidity at the same time. It was calculated as the percentage of time in which both temperature and relative humidity are inside the required range. Then the complementary percentage of time referred to the values out of this range was defined, specifying if the Shifted Index (SI) was related to temperature, or relative humidity, or both parameters together. In Fig. 18 even when the PI is higher than 90%, some values shift out of the correct range for preservation. Synthetically, the analysis shows that, considering the whole exhibit area, the PI values are far higher than a "warning value" of 90%, which means 10% of data outside the required ranges.

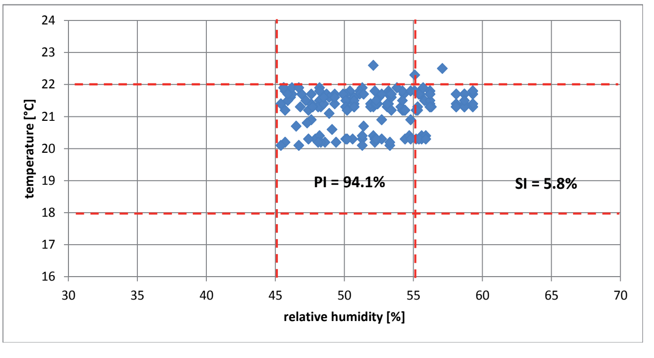


Figure 18. Global Performance Index on Temperature and Relative Humidity Matrix

Another important result regards CO_2 data. In Fig. 19 a temporal diagram of CO_2 behaviour inside and outside the showcases is shown. It is important to note how the showcases protected the artifacts thus reducing the CO_2 emissions under the alert threshold.

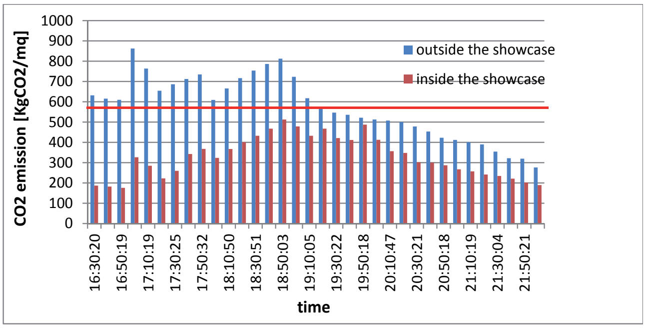


Figure 19. Temporal Diagram of CO_2 Emission

6. Conclusion

This chapter deals with the design, optimization and development of a practical solution for application to the Cultural Heritage monitoring and control. The overall system was addressed in terms of the experiences platform, network issues related both to the node's communication protocol and gateway operations up to the remote user's suitable interface. In particular, the proposed solution was installed in several museums and was used to monitor the art objects during their transport from a museum to another. The experimental results highlighted a noticeable performance as far as the data collecting reliability, the system robustness and the usability are involved. This allows the application of the solution under investigation to the more general field of environmental monitoring, due to its flexibility, scalability, adaptability and self-reconfigurability.

Acknowledgement

The authors would like to thank T.T. Tecnosistemi S.p.A.

Author details

Luca Bencini
T.T. Tecnosistemi S.p.A., Via Rimini 5, 59100 Prato, Italy

Stefano Maddio
Dept. of Electronics and Telecommunication, University of Florence, Via di Santa Marta 3, 50139 Florence, Italy

7. References

[1] Akyildiz, I.F.; Su, W.; Sankarasubramaniam, Y. & Cayirci, E. (2001). Wireless sensor networks: a survey, In: *Journal on Computer Network*, Vol. 38, pp. 393-422, December 2001.

[2] Yick, J.; Mukherjee, B. & Ghosal, D. (2008). Wireless sensor network survey, In: *Journal on Computer Network*, Vol. 52, pp. 2292-2330, April 2008.

[3] Sohraby, K.; Minoli, D. & Znat, T. (2007). Wireless Sensor Networks. Technology, Protocols and Applications, In: *John Wiley & Sons*, 2007.

[4] Karl, H. & Willing, A. (2005). Protocols and Architectures for Wireless Sensor Networks, In: *John Wiley & Sons*, 2005.

[5] Akyildiz, I.F.; Su, W.; Sankarasubramaniam, Y. & Cayirci, E. (2002). A Survey on Sensor Networks, In: *IEEE Communications Magazine*, Vol.40, pp. 102-114, August 2002.

[6] Al-Karaki, J.N. & Kamal, A.E. (2004). Routing Techniques in Wireless Sensor Networks: a Survey, In: *IEEE Communication Magazine*, Vol. 6, pp. 6-28, 2004.

[7] SENSIRION - SHT75. [Online: `http://www.sensirion.com`].

[8] MicroSemi - TAOS TSL252R-LF. [Online: `http://www.microsemi.com`].

[9] City Tech - EZT3NDH, EZT3SH, O33E1F. [Online: `http://www.citytech.com`].

[10] XSens - Mti. [Online: `http://www.xsens.com`].

[11] AK-NORD - XT-Nano-XXL Ethernet Module. [Online: `http://www.ak-nord.de`].

[12] Maddio, S., Cidronali, A. & Manes, G. (2011). A new design method for single-feed circular polarization microstrip antenna with an arbitrary impedance matching condition, In: *IEEE Transactions on Antennas and Propagation 59*, pp. 379-389, January 2011.

[13] Cidronali, A., Maddio, S., Giorgetti, G. & Manes, G. (2010). Analysis and performance of a smart antenna for 2.45-ghz single-anchor indoor positioning, In: *IEEE Transactions on Microwave Theory and Techniques, IEEE Transactions on 58*, pp. 21-31, January 2010.

[14] Sveda, M.; Benes, P; Vrba, R. & Zezulka, F. (2005). Introduction to Industrial Sensor Networking, In: *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, pp. 10-24, 2005.

[15] Shakkottai, S.; Rappaport, T.S. & Karlsson, P.C. (2003). Cross-layer Design for Wireless Networks, In: *IEEE Communication Magazine*, Vol. 41, pp. 74-80, October 2003.

[16] Jain, J.N. & Agrawala, A.K. (1992). Open Systems Interconnection: Its Architecture and Protocols, In: *Mcgraw-Hill Series on Computer Communications*, November 1992.

[17] Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), In: *IEEE Standard 802.15.4*, 2003.

[18] Ye, W.; Heidemann, J. & Estrin, D. (2002). An Energy-Efficient MAC Protocol for Wireless Sensor Networks, In: *Proc. of INFOCOM 2002*, Vol. 3, pp. 1567-1576, June 2002.

[19] Dam, T. & Langendoen, K. (2003). An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks, In: *Proc. of SENSYS 2003*, pp. 171-180, November 2003.

[20] Pecorella, T. NePSing - Network Protocol Simulator ng. [Online: `http://nepsing.sourceforge.net`].

Chapter 13

Application of Wireless Sensor Network for the Monitoring Systems of Vessels

Hussein Kdouh, Gheorghe Zaharia, Christian Brousseau,
Hanna Farhat, Guy Grunfelder and Ghaïs El Zein

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48276

1. Introduction

Wireless Sensor Networks (WSNs) have gained worldwide attention in recent years, particularly with the proliferation of Micro-Electro-Mechanical Systems (MEMS) technology which has facilitated the development of smart sensors. Smart sensors are small devices composed of one or more sensors, a memory, a processor, a power supply and a radio unit. They can sense the environment, measure and send data wirelessly to control unit for further processing and decisions. WSNs have great potential for many applications such as habitat monitoring (Polastre et al., 2004), intrusion detection and target tracking and surveillance (Arora et al., 2004), oceanography (Tateson et al., 2005), environmental monitoring (Barrenetxea et al., 2008a, 2008b; Padhy et al., 2005; Selavo et al., 2007), structural health monitoring (Paek et al., 2005), infrastructure monitoring (Stoianov et al., 2007), precision agriculture (Langendoen et al., 2006), biomedical health monitoring (Gao et al., 2005), and hazardous environment exploration and seismic sensing (Werner-Allen et al., 2006).

Structures, including bridges, buildings, dams, pipelines, aircraft, ships, among others, are complex engineered systems that ensure society's economic and industrial prosperity. Monitoring systems have been implemented for these structures to monitor their operation and behaviour against incidents. The monitoring system is primarily responsible for collecting the measurement output from sensors installed in the structure and storing the measurement data within a central data repository. To guarantee that measurement data are reliably collected, structural monitoring systems employ wires for communication between sensors and the repository. While wires provide a very reliable communication link, their installation in structures can be expensive and labour-intensive. With the emergence of wireless sensor technologies, industrial and academic groups have started to investigate the feasibility of WSN to replace the current wired monitoring systems (Lynch et al., 2006). Ships constitute an important part of modern systems widely used in armed conflicts and commercial purposes such as fishing and transporting passengers and cargos. Ships manufacturers and navy companies aim to use automation on board ships as much as possible in order to improve security and reduce the number of crew members. Modern ships are equipped with automatic monitoring systems which control and ensure the safety and accuracy of the whole ship operation. Current shipboard monitoring systems use extensive lengths of cables to connect several thousands of sensors to central control units. Tens of kilometres of cables may be installed on board a ferry-boat, increasing its cost, weight and architecture complexity. In addition to the high cost of wires installation during ships construction, vessels represent a complex and harsh environment in which extensive lengths of wires are vulnerable to detriments such as heat, moisture and toxic agents. Hence, using wireless communication between sensors and control units on board ships presents several advantages over wired solution. Radio waves travel through space, i.e. the additional cost, weight and complexity produced by the routing of cables through the structure of a vessel, are eliminated. Moreover, wireless systems are easily and inexpensively reconfigured. Therefore, using the WSN technology for shipboard monitoring systems can be a cost-effective and survivable solution. Wireless sensor nodes are capable to form a large scale (up to thousands), self-organising and self configurable ad hoc network with low cost and low power consumption devices.

However, electromagnetic waves propagation on board a vessel is a serious challenge. Several factors decrease the performance of wireless networks in this particular environment. Metallic bulkheads, made often of steel, can severely decrease the power of received signals. Moreover, multipath effects leading to multiple delayed copies of the transmitted signal at the receiver may also decrease the radio communication data rate. A propagation study must be carried out in this harsh environment to ensure the reliability of radio links and the WSN feasibility.

This chapter studies the feasibility of WSN on board ships. Several measurement campaigns are conducted on board a ferry-boat to verify the possibility of wireless communications between ship parts and to analyse the performance of WSN on board. These measurements aim at determining path loss models for typical shipboard environments and testing the possibility of wireless communication between adjacent rooms or adjacent decks. Using the results of these experiments, a WSN is tested on board the ferry. The results obtained from the measurement campaigns are then used to propose an architecture for a large-scale shipboard WSN. As the network test uses a limited number of nodes, the full monitoring system based on the proposed architecture is simulated using a network simulator.

2. Related works

Several research teams have investigated the possibility of using of wireless sensors on board vessels.

In (Mokole et al., 2000), a feasibility study of wireless communications using Commercial Off-The-Shelf (COTS) wireless modems that communicate at radio frequencies from 800

MHz up to 3 GHz was conducted on board vessels. Authors have verified that radio communications are possible between adjacent rooms even when watertight doors are closed.

In (Estes et al., 2001), measurement campaigns were carried out on board various naval vessels to verify the feasibility of intra- and inter-compartment radio communications. The measurement results have shown that ship bulkheads severely decrease the power of received signals of about 20–30 dB but communication through two or three bulkheads is found to be still possible. They explain this result by the presence of a number of non-steel elements in the bulkheads (e.g. hatch seals, ducts, cable transits) that allow radio signals to penetrate.

In (Schwartz, 2002), a new shipboard monitoring system using wireless sensors interfacing to a ship Local Area Network (LAN) through 802.11 Wireless Access Points (WAPs) was proposed. The system has been validated successfully on numerous naval vessels including the USS Monterey and the ex-USS Shadwell.

Authors in (Brown et al., 2003) presented a process template to assist the information and process control technologist in successfully deploying today's COTS WLAN systems. The process focuses on an eight-step process that balances analytical modelling requirements with empirical surveys to qualify below deck noise, signal propagation and realistic connectivity expectations.

Authors in (Ploeger et al., 2003) proposed a wireless shipboard monitoring system constituted of wireless data acquisition nodes, called Intelligent Components Health Monitor (ICHM), that are capable to collect sensor data from analog sensors and communicate these data via Bluetooth wireless radios to a centralized data repository, called Compartment Health Monitor (CHM).

Authors in (Li et al., 2003; Ou & Li, 2003) studied the feasibility of using wireless sensors for monitoring the health of offshore oil platforms. The proposed WSN is constituted of multiple sensor nodes wirelessly connected to a base station which collects the data for processing and distribution through a LAN or the Internet.

(Takahashi, 2004) reported on the use of wireless sensors for wireless monitoring of oil tankers. Wireless sensors manufactured by Dust Networks are being installed throughout various oil tankers, especially in critical regions where structural or mechanical problems could potentially occur.

Authors in (Krishnamurthy et al., 2005) focused on the preventive equipment maintenance in which vibrations signatures are gathered to predict equipment failure. Based on application requirements and site surveys, they have proposed and tested an architecture for this type of application on board an oil tanker in the North Sea. The sensor network including 150 accelerometers, 26 sensor nodes, 4 Stargates and 1 PC has been deployed and tested during four months on board the ship.

Authors in (Park et al., 2008) carried out some experiments using ZigBee devices on board a ship. Their communication tests have shown that intra-compartment wireless communications are possible and inter-compartments wireless communications are almost impossible. Based on these results, they have successfully tested a hybrid WSN using ZigBee for intra-compartment communications and Power Line communications (PLC) for inter-compartments communications.

Moreover, authors in (Paik et al., 2009) carried out some transmission tests using two ZigBee protocol analyzers to evaluate the performance of wireless communications on the passenger deck of a ship. Four scenarios including communication between a cabin and the corridor, in the corridor and between adjacent decks with and without entrance door closure, have been considered. In addition, a ZigBee-based WSN has been successfully tested in the engine room of the ship.

Authors in (Pilsak et al., 2009) investigated the propagation conditions of 2.4 GHz RF waves on a bridge of a modern cruise vessel which is important for evaluation of the ElectroMagnetic Compatibility (EMC) behaviour of the electronic bridge equipment. The intention of such an evaluation is to ensure that electronic equipment, as well as the wireless transmission line, is not disturbed. The bridge has been simulated with a 3D model which includes the material data of the different objects on the bridge. A ray tracing algorithm has been applied to this model and the maximum data rate of a 2.4 GHz wireless LAN system has been simulated. In addition, measurements on the bridge have been performed to back up the simulation results and to investigate the real case.

Authors in (Kang et al., 2011) proposed a new method of tracking the crew member location using ZigBee tags and routers. Their method was tested and proved its viability on board steel-structured ships. The authors think that this method may assist the onboard training organizer and commanding officer by providing complete information to base its decisions.

Finally, authors in (Kdouh et al., 2011a, 2011b, 2011c, 2012) reported on the feasibility of WSN on board ships. Several measurement campaigns have been conducted on board several ferries to verify the possibility of intra-, inter-compartments and inter-decks radio communication. A WSN has been tested successfully on board a ferry. The obtained results of these works will be detailed in the remaining of this chapter.

3. Measurement sites

'Acadie' is the ship used for this study. It is a ferry boat from the 'Compagnie Océane'. The 'Acadie' is constituted of the following decks, arranged vertically from bottom to top: the bottom deck which houses the main engine room, the control room and the crew's cabins; the main deck which is a parking; the passenger deck, and the bridge deck which contains the wheel house. Four typical environments are considered for the propagation measurements: the engine room, the parking, the passenger deck and the crew's cabins.

The engine room of 'Acadie' is composed of the main engine room and the control room. These two rooms are separated by a bulkhead and a watertight door which have both a big glass window. The engine room contains engines, pumps, generators and valves. The other part of the bottom deck houses the crew's cabins. This part is separated from the engine room by a thick metallic bulkhead. The cabins doors are made of wood.

The parking of 'Acadie' is constituted of a big hall with metallic walls including some glass windows and some small rooms (in the front section) with metallic watertight doors. Measurements were carried out on board the ferry when it was moored to the harbour. There were no vehicles parked in the parking. The parking is connected to upper and lower decks by stairways that have a metallic watertight door on the parking side.

The passenger deck of 'Acadie' is a big hall with metallic walls including glass windows. It is composed of passengers' seats and tables. This environment is composite and constituted of several types of materials such as wood, glass and steel.

4. Propagation measurements

This section describes the propagation measurement campaign conducted on board 'Acadie'. It includes the measurement procedure, results and analysis.

4.1. Measurement procedure

Due to the low data rate of a shipboard WSN, Continuous Wave (CW) measurements are sufficient to characterize the propagation effects related to a WSN deployment because the bandwidth of the transmitted signal is much less than the coherence bandwidth of the propagation channel. The transmission system is composed of a signal generator, an omnidirectional conical monopole antenna and some connecting cables. The signal generator delivers 0 dBm sinusoidal signal at a frequency of 2.45 GHz (ISM radio band - Industrial Scientific and Medical). This ISM frequency band has been selected as it is used by most existing standards dedicated to WSN (Yick et al., 2008).The receiver is composed of a spectrum analyzer operating in a zero-span mode, a laptop to collect and save measurements data, an antenna positioner and connecting cables.

Each shipboard environment was measured using a standard procedure. The transmitting (Tx) antenna, which has a height of 1.80 m, is placed at a fixed location. Path loss measurements are performed using a receiver (Rx) with a 1.80 m antenna height. The receiver is placed at different locations in each shipboard environment. Tx and Rx locations are marked on a digital map to calculate the Tx-Rx separation distance. These experiments rely on narrowband measurements of a CW signal at 2.45 GHz performed to determine the path loss. The received power varies over a small area due to multipath-induced fading. However, averaging the received power values along 20 wavelength circular track using 250 power samples, yields a reliable estimation of the local average power independent of signal bandwidth (Durgin et al., 1998). The average of the received power values in Watts is used for all path loss estimations.

4.2. Measurement scenarios

Fig. 1 shows the transmitter locations (Tx1 to Tx4), the receiver locations (blue squares), the layout of the ship and the measured path loss for all environments considered on board 'Acadie'. In the passenger deck, the transmitter was placed at the Tx1 location and the receiver was placed at 16 different locations. In the parking, the transmitter was placed at the Tx2 location and the receiver was placed at 21 different locations. In the engine room, the transmitter was placed in the control room (Tx3 location) and the receiver was placed at 14 different locations in the main engine room. To characterize the communication between decks, the transmitter was placed at the location Tx4 in the parking (2 m in front of the watertight door) and the receiver was placed at 11 different locations in the crew cabins. These two decks are connected by metallic stairs. The entrance watertight door to the stairway in the parking was closed during these experiments. The other three stairways connecting the parking to the engine room and the passenger deck have the same architecture. The results of this experiment can be generalized to characterize the communication between decks.

4.3. Results analysis

The main configurations of communication between nodes in a future shipboard WSN are:

- communication between nodes placed in the same room
- communication between nodes placed in different rooms
- communication between nodes placed in different decks

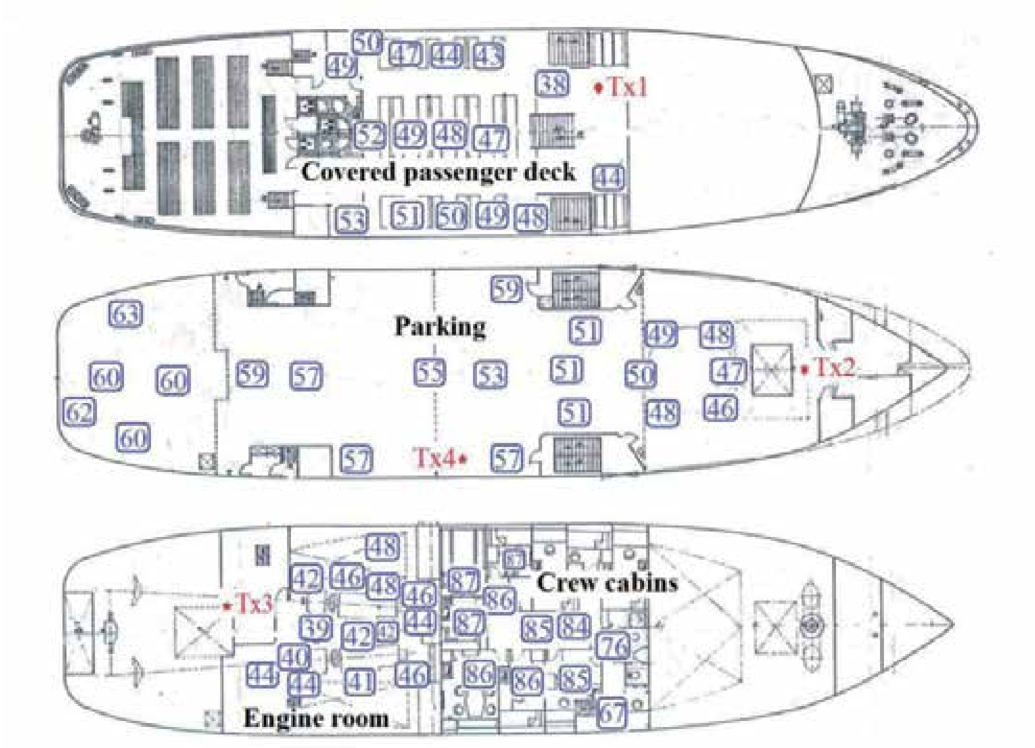


Figure 1. Layout of different parts of the 'Acadie' vessel, and locations of the transmitter Tx1, Tx2, Tx3 and Tx4 (in red), and the receivers (blue squares). Values in the blue squares are the path loss in dB

A communication is considered as possible when the received power is higher than -85 dBm. This threshold is related to the receiving sensitivity of sensor nodes that will be used later in the WSN experiment (Memsic Technology, 2007a).

4.3.1. Communication between nodes within the same room

The three considered environments in this case are: the engine room, the parking and the passenger deck. Measurement results are used to determine the relation between the path loss and the distance between nodes in each environment. Average path loss for a separation distance d between the transmitter and the receiver is expressed as a function of distance by using the following expression (Rappaport, 2002):

$$\overline{PL(d)} = \overline{PL(d_0)} + 10n\log_{10}(d / d_0) \tag{1}$$

where n is the path loss exponent which indicates the rate at which the path loss increases with distance and d_0 = 1 m is the reference distance. This model does not consider different surrounding configurations for the same Tx-Rx separation distance d. Measurements have shown that at any value of d, the path loss PL(d) for a particular location is random and has a log-normal distribution around its mean distance-dependant value. Hence, path loss can be expressed as:

$$PL(d) = \overline{PL(d_0)} + 10n\log_{10}(d / d_0) + X_\sigma \tag{2}$$

where X_σ is a zero-mean Gaussian distributed random variable (in dB) with standard deviation σ (also in dB). The log-normal distribution describes random effects of shadowing or multipath propagation which occur over a large number of measurement locations having the same separation distance but with different levels of clutter on the propagation paths (Rappaport, 2002).

The results of measurements performed on board the 'Acadie' vessel have shown a significant correlation with model (1). Fig. 2 shows path loss values as a function of distance for all environments. Shadowing effects have been taken into account by the Gaussian distributed random variable with σ computed as the standard deviation of the error between the measurements and the model (1) results.

The values of $\overline{PL(d0)}$, n, and σ have been computed from measured data using linear regression (Minimum Mean Square Error MMSE estimation). The parameters obtained for the three environments are given in Table 1 where ρ is the correlation coefficient between measurements and model results. The large values of ρ show a significant correlation between measurement results and the path loss model. Nevertheless, the value of ρ in the engine room is lower than that in other environments. This difference may be explained by the complex arrangement of metallic machines and tubes in this environment, which randomly scatters, reflects and diffracts the radio waves. The arrangement is more homogenous in the passenger deck and the parking.

Some preliminary conclusions may be drawn from the values of n. The path loss exponent is equal to 1 in the engine room of 'Acadie'. This result can be explained by the presence of metallic walls and ceiling and the absence of significant radio leakage between the engine room and the neighbourhood (the access between the engine room and the parking was closed during measurements). The transmitted energy is then kept within the engine room. The engine room is then similar to a reverberant chamber. Moreover, the path loss exponent in the parking is equal to 1.61 which is lower than the free space path loss exponent. This result is explained by the guiding effect of metallic walls and ceiling. However, the difference between the engine room and the parking exponents is explained by the presence of glass windows in the parking walls which allow EM leakage for radio waves. The transmitted energy is not kept inside the parking like in the engine room where the walls are completely metallic. Furniture obstructing the visibility between Tx and Rx explains the larger value of n in the covered passenger deck.

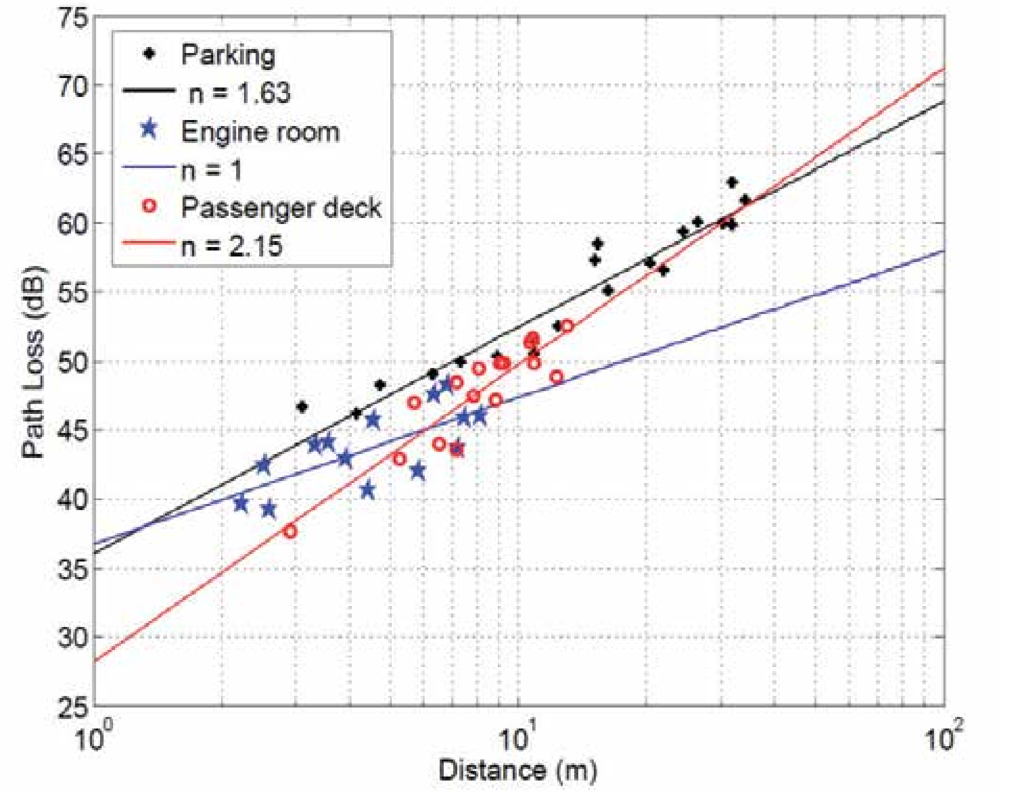


Figure 2. Scatter plot of path loss versus Tx-Rx distance within the same room

Environment	n	$\overline{PL(d_0)}$ (dB)	σ(dB)	ϱ
Engine room	1	36.76	1.37	0.72
Parking	1.63	36.10	1.21	0.96
Passenger deck	2.15	28.19	1.25	0.90

Table 1. Path loss parameters

4.3.2. Communication between nodes placed in different rooms

The second studied configuration is the communication between nodes placed in different rooms of the same deck. EM waves propagation is considered in the bottom deck and the parking which contain several rooms. In this case, the propagation path between Tx and Rx is obstructed by bulkheads and doors.

The first scenario is the communication between the crew's cabins and the engine room. As stated before, these two parts are separated by a thick totally metallic bulkhead. The transmitter is located in the corridor between crew cabins and the receiver is moved in the engine room. No signal has been received, in spite of the small Tx-Rx separation distance. This is explained by the huge attenuation of the thick metallic bulkhead and the absence of openings allowing EM leakage between these two adjacent areas.

The second scenario is the communication between nodes located in two adjacent rooms with a common door. Two types of doors may be considered on board 'Acadie': the metallic watertight doors that are mainly used at the entrance of stairways connecting the parking to other decks, and between small rooms located in the front section of the parking; and the wooden doors of the crew's cabins. Several experiments have been conducted to determine the excess path loss due to closing a door between two nodes, using the following experimental protocol. Tx and Rx are located in the two sides of the door and path loss is measured when the door is opened and when it is closed (with the same locations of Tx and Rx for both cases). Excess path loss due to the door closure is determined as the difference between the two measured values. The results have shown that the closure of a metallic watertight door decreases the received signal by an average value of 20 dB (with a standard deviation of 3 dB). However, the effect of wooden doors was negligible (no more than 0.5 dB of attenuation).

Several conclusions may be drawn from these experiments regarding the configuration between two nodes placed in adjacent rooms. If the common bulkhead between rooms is totally metallic and does not support a door, the communication is impossible. Otherwise, in the presence of a door, the communication is always possible (door opened or closed). Closing a watertight door on the propagation path between nodes decreases the transmitted signal level by up to 25 dB. However, the presence of the two closed watertight doors between two nodes makes their connectivity impossible.

4.3.3. Communication between nodes in different decks

Path loss levels of measurements between the parking and the passenger deck (Fig. 1) show that the transmitter located in front of the watertight door in the parking is not able to cover the total area of the crew's cabins deck. The maximum acceptable path loss is 85 dB, which is less than most of the values found in this deck. The variation of path loss values in this configuration does not depend directly of the Tx-Rx separation distance. It depends on the closeness of the Rx and Tx to the stairway. This variation indicates that stairways are the main sources of EM leakage between adjacent decks. Hence, placing intermediate sensor nodes in the stairways is necessary to maintain the connectivity of shipboard WSN.

5. Wireless sensor network test

This section describes the deployment of a WSN based on the conclusions drawn from the propagation study. Firstly, the technology used in the experiment is described and then, in the second part, the deployment procedure is presented. Finally, the obtained results are presented and discussed.

5.1. Technology used for WSN test

The shipboard WSN test was carried out using Crossbow's MICAz wireless sensor nodes (motes) (Memsic Technology, 2007a). MicaZ, which is IEEE 802.15.4 compliant, is a tiny wireless measurement system designed specifically for deeply embedded sensor networks. Each node is composed of a processor, an internal memory, a 2.4 GHz radio transceiver, two 2A batteries and a sensor board. It has a maximum data rate of 250 kbps. Embedded sensors can measure temperature, humidity, barometric pressure, ambient light and acceleration. The Crossbow's XMesh routing protocol, which is a link-quality based dynamic routing protocol that uses periodic route update messages from each node to estimate link quality, has been used in this experiment. Each node listens to the radio traffic in the neighborhood and selects the parent that would be the least costly in terms of transmissions number to reach the base station (Memsic Technology, 2007b). The network is composed of 12 sensor nodes and one gateway connected to a laptop via a USB cable. The laptop runs the MoteView 2.0 software which is a graphical user interface that allows visualizing the real time data sent by the WSN to the base station and the network topology evolution during the test.

5.2. Deployment procedure

The choice of the locations of nodes is based on the results obtained from the propagation study. Previous results have shown that EM waves propagation is possible between decks through stairways. To ensure the connectivity between the four decks of the shipboard WSN, relay nodes are first installed in the stairways. Hence, the deployment procedure has continued by installing the following nodes in stairways (Fig. 3):

- Node 3 between the crew's deck and the parking (the watertight door is closed)
- Node 2 between the engine room and the parking (the watertight door is open)
- Node 7 between the parking and the passenger deck (the watertight door is open)
- Node 11 between the covered passenger deck and the non-covered passenger deck (this stairway has a wooden door which was closed during the test).

The base station is installed in the control room (same location of Tx3 in Fig. 1). Node 1 is installed on one of the two main engines in the engine room, node 4 is installed in the crew's deck and node 9, in the covered passenger deck. Node's installation is different in the parking where several cases have been distinguished as a function of the number of watertight doors between the transmitter and the receiver. Node 8 is installed in the small room located in the front section of the parking. The watertight door of this room is closed during the test. Node 6 is installed in the middle of the parking and node 5 is installed in front of the second stairway located between the engine room and the parking. Node 12 is used to collect data from the wheel house. There are no stairs between the wheel house and the lower decks. Thus, node 10 is installed on the bridge deck as an intermediate node between node 12 located in the wheel house and node 11 located in the stairways between the covered and the non covered passenger deck.

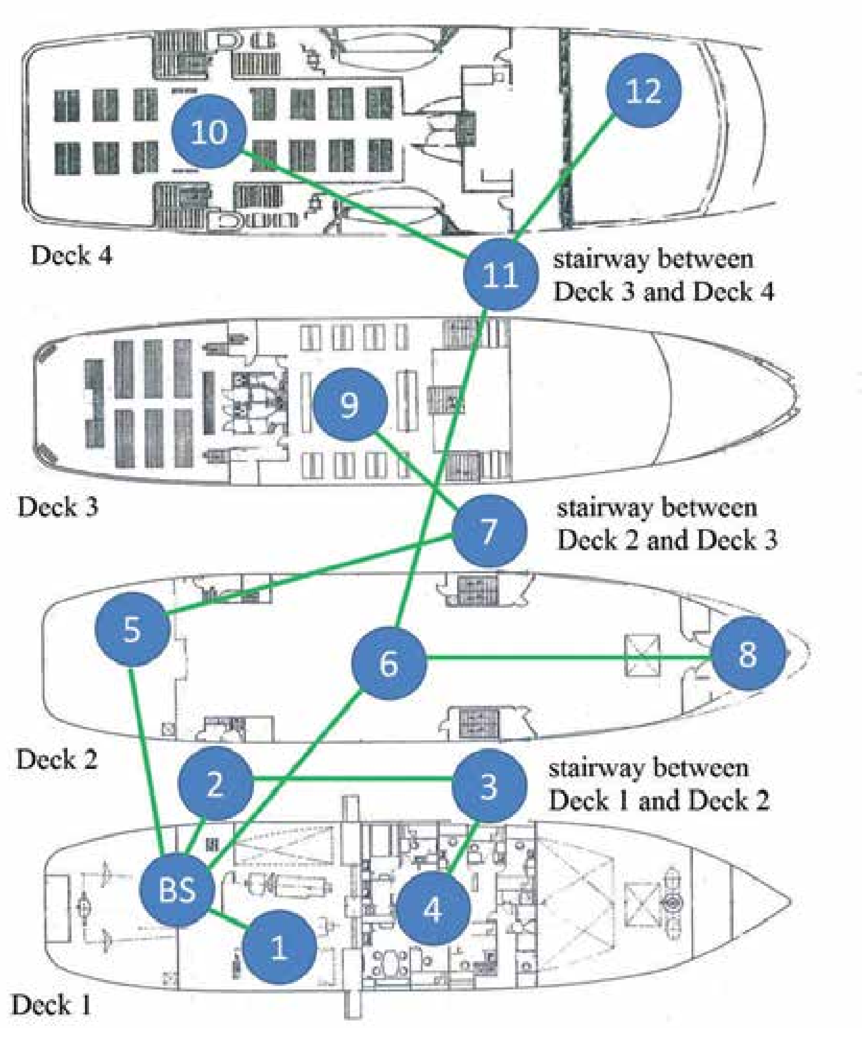


Figure 3. Locations of sensor nodes on board 'Acadie'

5.3. Network results

Analysis of the performance of the network has begun with the statistics of the packets sent by all the nodes during the experiment. Fig. 4 presents the percentage of originated, forwarded and dropped packets of the 12 nodes in the WSN. 'Originated' packets include all data, node health, neighbour health and route update packets originated at the node. 'Forwarded' packets are the packets that the node has received from other nodes and forwarded to other nodes. 'Dropped' packets are the packets that the node has dropped. Packets are considered dropped when 1 packet has been retransmitted 8 times without receiving the link-level acknowledgement.

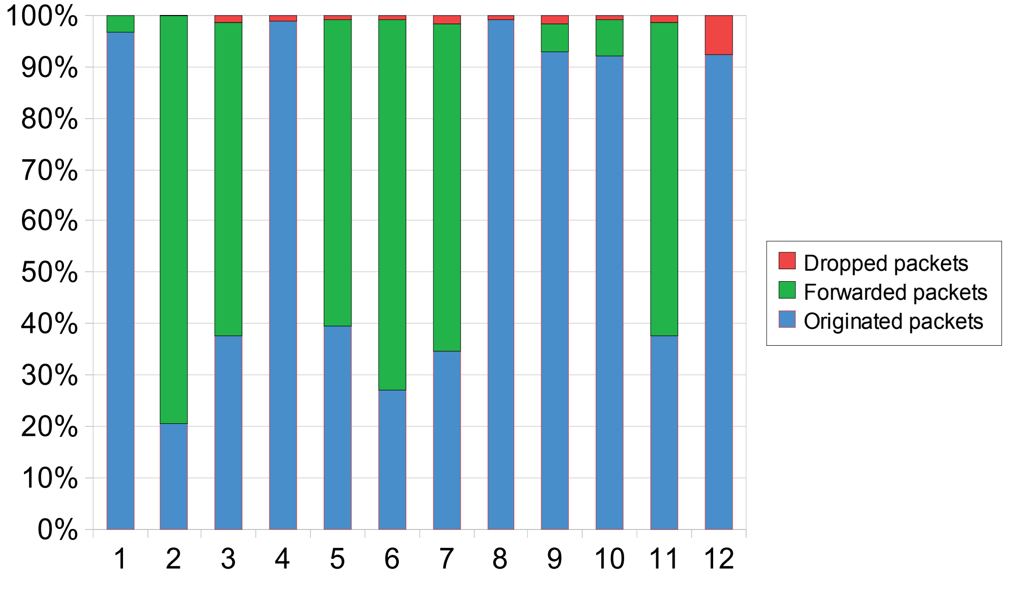


Figure 4. Percentage of dropped, forwarded and originated of all sensor nodes

The obtained results (Fig. 4) show that less than 2 % of packets have been dropped for most of nodes (only node 12 has 7 % of dropped packets due to its particular location in the wheel house, separated from other ship parts). The small percentage of dropped packets reflects significant efficiency of the XMesh routing protocol in such hostile environments. It can be also noticed that the huge amount of forwarded packets comes from the nodes 2, 3, 5, 6, 7 and 11, due to the location of these nodes in the stairways. The whole network connectivity is based on the convenient location of these nodes. Nodes in upper decks route their data mainly through stairway nodes, as the radio signal penetration is impossible through metallic ceiling and floors.

In order to improve knowledge about radio propagation inside the vessel, the paths followed by packets from source nodes towards the base station have been studied. As previously stated, the sensor nodes are pre-programmed by the Crossbow XMesh routing protocol. Therefore, a sensor node selects the next hop which minimizes the number of transmissions required to send a packet to the base station. The selected parent node (next hop node) is then characterized by its closeness to the base station (in term of number of hops) and its good link quality with the source node. Hence, the choice of the next hop may be an indicator of the quality of links between a sensor node and its one-hop neighbours. Table 2 shows the parent nodes of each sensor node and the percentage for each one during the test.

NodeID	GW	1	2	3	4	5	6	7	8	9	10	11	12
1	100												
2	100												
3			56			34	9	1					
4				100									
5	65		35										
6	65		35										
7			33	10		42	15						
8						1	99						
9							45	43				12	
10								49				51	
11							70	27		3			
12											18	82	

Table 2. Percentage of selected parents for each sensor node of the network

Several remarks may be drawn from this table. It can be noticed that node 8 has never selected node 3 (or node 7) as parent node despite the small distance between them. This behaviour is in agreement with the statement that two nodes separated by two closed watertight doors cannot be connected in the parking (nodes 3 and 7 are located in two stairways with closed doors). However, this connection remains possible when only one closed watertight door separates the two nodes (which is the case of nodes 6 and 8). Furthermore, it can be noticed that nodes 1 and 2 were directly connected to the base station (GW = Gateway) during the test. This was expected, as these three nodes are located in the engine room where the probability of outage between two nodes is very low (due to the low path loss exponent). Node 4 has always node 3 as parent node. Connection between the crew's deck and the parking is not possible because the watertight door in the entrance of the parking is closed. In spite of the small distance between nodes 4 and 1, the connection between them is impossible since the engine room and the crew's deck are separated by a thick metallic bulkhead. However, it can be noticed that node 6, located in the middle of the parking, is directly connected to the base station GW for 65 % of of the forwarded packets (as well as node 5) and node 9 is connected to node 6 for 45% of the time. This can be explained by the fact that the two watertight doors of the first stairway between the engine room and the parking and the first stairway between the parking and the passenger deck are opened. These two nodes used the intermediate nodes located in the stairways (nodes 7 and 2) for the remaining time when the direct connection becomes impossible. Finally, node 12 located in the wheel house has node 11 (82 %) and node 10 (18 %) as parent nodes. The direct connection between nodes 12 and 11 is probably provided by the signal reflection on the metallic tour upside the non-covered passenger deck.

6. Hierarchical architecture for large-scale shipboard WSN

The following concluding remarks can be drawn from the measurement campaigns:

- Ships (especially ferry-type) are built of metallic blocks that constitute decks and rooms.
- Wireless communications between adjacent rooms are possible in the presence of non-conductive materials in the common bulkhead.
- Watertight doors are the main source of radio leakage between adjacent rooms. Closing a watertight door induces an attenuation up to 25 dB.
- Stairways are the main source of radio leakage between adjacent decks.
- Wireless communication between spaced nodes is possible through multi-hop communications.

These conclusions are used in this section to propose an architecture for a large-scale shipboard WSN.

6.1. Proposed architecture

As previously stated, the shipboard monitoring system may contain several thousands of sensors located in all compartments. Some rooms, such as the engine rooms, may contain hundreds of sensors. Using the previously stated concluding remarks, a hierarchical WSN architecture adapted to the particular characteristics of the shipboard environments is proposed. In this architecture, the network will be divided into groups and different nodes levels are defined, based on the functions and resources of nodes. The radio propagation study has shown that the metallic structure of ships makes each room (which is similar to a metallic cube) quasi isolated (from a wireless propagation point of view). Therefore, it has been decided to divide the network into zones where each metallic room is a zone. Three types of nodes may be found in this architecture: Sensor Nodes (SN) which collect sensing data from the environment, Border Nodes (BN) which collect data from SNs, and Gateway Nodes (GN) which collect data from the BNs and send them through a wired connection to the central processing units. Two types of wireless communications are distinguished: the intra-zone communications and the inter-zone communications.

6.1.1. Sensor nodes

This level is constituted of SNs distributed in all ship rooms. Different data may be measured by these nodes such as temperature, pressure, humidity, fire, tank level, water level, etc. depending on the application. One SN may be connected to several sensors if their locations are close (case of the engine room where hundreds of sensors are located in a small area). If SNs are powered by batteries, their power consumption must be optimized. As the radio unit (Tx and Rx) consumes the most of the energy, it must be in the sleep mode as much as possible. Therefore, the number of transmissions must be optimized. In the confined metallic rooms, one-hop communication is sufficient between any nodes placed in the same room. Sensor nodes will not be intended to forward data from other nodes, which can greatly reduce their power consumption. Radio units are then turned on only when sensor nodes want to send their sensing data to the border node. These data may be periodic or event driven. In order to minimize the number of transmissions, a Hard Threshold (HT) and a Soft Threshold (ST) may be predefined for each application. It is not necessary that a SN sends its data continuously to its BN. Instead, it saves the last sent data and continues to sense its environment. Measured values will be compared firstly to HT. If it exceeds this value (higher or lower depending on the application), the data will be sent. If not, the difference between the last value and the measured value will be compared to ST. If the difference exceeds ST, the value will be sent. This procedure reduces the number of transmissions to only urgent cases (exceeding HT) or to important value changes (exceeding ST). A careful attention must be given to the Medium Access Control (MAC) layer in order to minimize collisions. As the IEEE 802.15.4 standard is adopted for this study, the used MAC algorithm is CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). Another contention free mechanism is possible in this standard for critical applications.

6.1.2. Border nodes

Border Nodes (BNs) are the second level of the proposed architecture. Each BN is responsible of all or a part of the sensor nodes in a metallic room. BNs are placed in front of doors borders of each room. More than one BN may be placed in a room if it has several doors, giving multiple choices for SNs to join the network. SNs send their data to BNs via one-hop intra-zone communication. BNs query and gather sensed measurements from SNs, and aggregate collected data (eliminate redundancy) before sending to the base station via multi-hop inter-zone communication. Different routing protocols may be adopted for inter-zone communication. Regarding the critical role of a BN (it is responsible of a cluster of sensor nodes), it must be always powered on. BNs may be powered by the mains supply of the ship. Therefore, the inter-zone routing protocol does not have to optimize the energy consumption of these nodes. Instead, the link quality and the number of hops to the base station must be optimized. XMesh, the used protocol in the network test, or other routing protocols existing in the literature, may be used for inter-zone communications.

6.1.3. Gateway nodes and central data repository

Gateway Nodes (GNs) aggregate data from the network, interface to the host, the Ethernet or the Internet (through satellites connections). Gateways form bridges to send and receive data between the central repository and the sensor network. Similarly to BNs, gateways play a vital role in the network. Hence, they are always powered by the mains supply of the ship. Depending on the network size on board the ship and the technology adopted, one or more gateways may be used. In case of multiple gateways, each gateway will form a sub-network using a frequency sub-band and all gateways will be connected to an Ethernet installed on the ship. This mechanism increases the network scalability and decreases the collisions rate.

Data aggregated by the gateways are sent to a central repository located usually in the control room or in the wheel house of the ship. Data are analyzed and conclusions concerning the current state of each room are drawn. Central data repository is equipped with a user visualization software and a graphical interface for managing the network and showing measured data.

6.2. Performance analysis

This section presents the performance analysis of the proposed architecture. It includes the network simulator description, the used standard and the simulation scenarios.

6.2.1. Network simulator

OPNET Modeler 16.0 (OPNET Technologies, n.d.) is used to simulate and evaluate the performance of the proposed shipboard WSN architecture. OPNET is a discrete-event and object-oriented simulator. Strength of OPNET in wireless network simulations is the accurate modelling of the radio propagation. Different characteristics of physical-link transceivers, antennas and antenna patterns are modeled in detail. In OPNET, the possibility of wireless link between a transmitter and a receiver depends on many physical characteristics of the involved components, as well as time varying parameters, which are modeled in the Transceiver Pipeline Stages. Parameters such as frequency band, modulation type, transmitter power, distance and antenna pattern are common factors that determine whether a wireless link exists at a particular time or can ever exist.

However, OPNET does not take into account the physical obstacles between Tx and Rx in indoor environments. Studying the performance of the shipboard WSN architecture must be preceded by a realistic modelling of the shipboard environments. Therefore, several objects and functions have been developed in the simulator to take into account the propagation challenges. Firstly, the log-normal path loss model determined from the propagation measurement campaign is not supported by the “Terrain Modeling” module of OPNET. Therefore, this model has been integrated in the “Received Power Pipeline Stage”. The parameters of the model depend on the Tx and Rx locations. Secondly, a wall object has been developed to simulate the ship bulkheads. A “path loss” attribute has been given to each wall to indicate its structure (totally metallic, metallic with openings, wooden wall, etc). The excess path loss due to the existence of a wall between Tx and Rx is also taken into account when determining the path loss in the “Received Power Pipeline Stage”. Finally, the ship has been modelled using its real dimensions.

6.2.2. ZigBee standard

ZigBee (ZigBee Alliance, 2005) is one of the most used standards for WSNs. It is based on the IEEE 802.15.4 standard with a theoretical transmission data rate equal to 250 kbps in a wireless link. ZigBee defines three types of nodes: end devices, routers and coordinators. The coordinator creates the network, exchanges the parameters used by the other nodes to communicate, relays packets received from remote nodes towards the correct destination, and collects data from the sensors. Only a single coordinator can be used in a network. A router relays the received packets and the control messages, manages the routing tables and can also collect data from a sensor. Routers and coordinators are referred to as Full Function Devices (FFDs). On the other hand, end devices, also referred to as Reduced Function Devices (RFDs), can act only as remote peripherals, which collect values from sensors and send them to the coordinator or other remote nodes. However, RFDs are not involved in network management, and therefore, cannot send or relay control messages.

According to the ZigBee standard, three different kinds of network topologies are possible: star, cluster-tree, and mesh. In a star network, there are a coordinator and one or more RFDs (end nodes) or FFDs (routers) which send messages directly to the coordinator (up to 65536 RFDs or FFDs). In a cluster-tree topology, instead, there are a coordinator which acts as a root and either RFDs or routers connected to it, in order to increase the network dimension. The RFDs can only be the leaves of the tree, whereas the routers can also act as branches. In a mesh network, any source node can talk directly to any destination. The routers and the coordinator, in fact, are connected to each other, within their transmission ranges, in order to facilitate packet routing. The radio receivers at the coordinator and routers must be "on" all the time. In the mesh network, the ZigBee standard employs a simplified version of the Ad-hoc On-demand Distance Vector (AODV) routing protocol (Perkins et al., 1999).

Due to previous features, the ZigBee standard has been chosen to test the proposed architecture. SNs will be formed by ZigBee end devices, BNs will be ZigBee routers and the GN will be a ZigBee coordinator. As it is impossible to cover all the ship by a star topology (due to metallic obstacles), mesh and tree topologies have been only considered.

6.2.3. Simulation scenarios

Table 3 summarizes the parameters used for simulation.

Parameter	Value
Maximum number of router or end devices per router	200
Route discovery timeout (s)	10
Maximum depth	10
Acknowledge wait duration (s)	0.05
Minimum value of the back-off exponent in the CSMA/CA	3
Maximum number of back-offs	4
Channel sense duration (s)	0.1
Data rate (kbps)	250
Receiver sensitivity (dBm)	-95
Frequency band (GHz)	2.4
Transmission power (W)	0.001
Packet inter-arrival time (s)	1
Packet size (bits)	120

Table 3. Simulation parameters

The sensor nodes have been deployed on the simulation model of the four decks of the ship as shown in Fig. 5. The network is constituted of 100 sensor nodes (routers and end devices) and one coordinator located in the bottom deck. As previously stated, in each room where end devices are located, routers have been placed in front of watertight doors and windows. The number of sensor nodes in each room is related to the real placement of sensors in the current monitoring system, which contains hundreds of sensors. The engine room (bottom deck) contains 150 sensors. The packets size sent by each sensor is 2 bytes. As the rooms on board ships are not large, it would be possible to connect several sensors to one node. It is supposed that each sensor node is equipped with 5 sensors (similar to MicaZ nodes used in the measurement campaign). Hence, the data packet size is equal to 120 bits (8 bits for the sensor ID and 16 bits for the measured data). Therefore, this scenario simulates a WSN with 500 sensors.

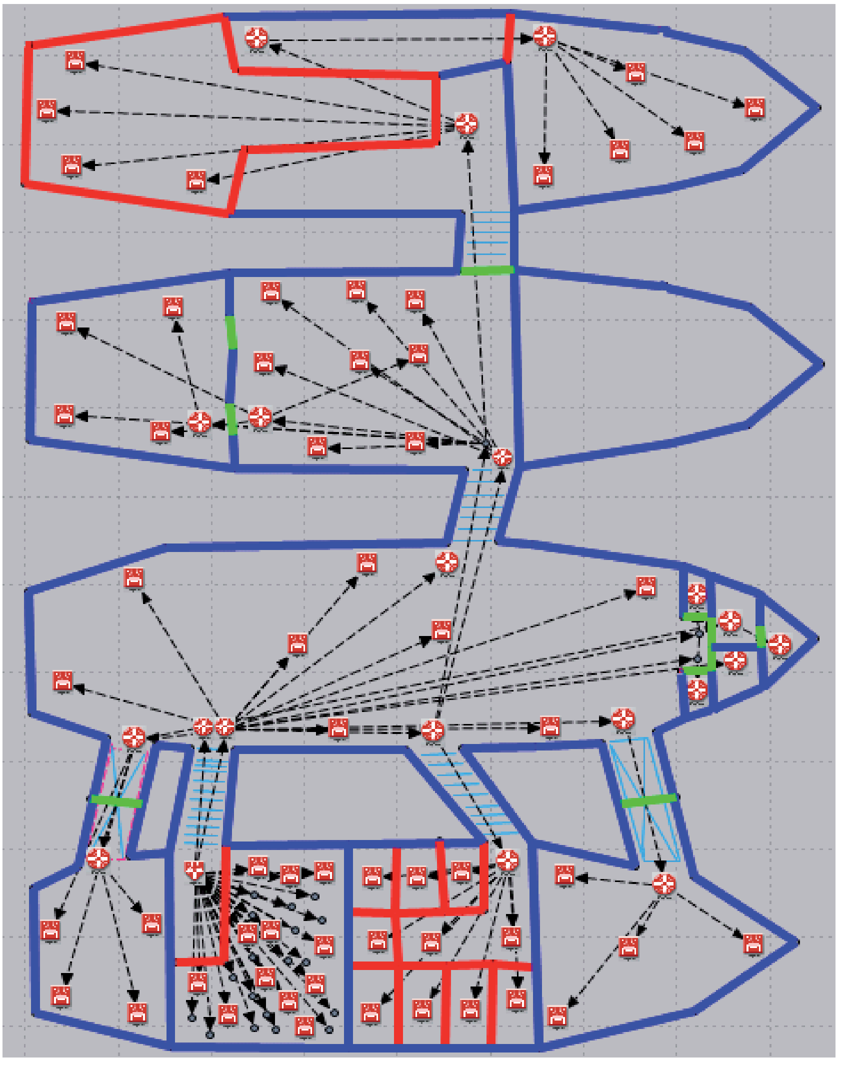


Figure 5. Layout of simulation model of Acadie and ZigBee WSN topology

6.2.4. Results and analysis

The objective of this study is to propose a reliable shipboard monitoring system based on wireless technologies. In spite of the important reduction of cost and complexity, this solution must provide a Quality-of-Service (QoS) similar to that provided by the current wired system. A monitoring system has hard requirements in terms of reliability and delays. All critical sensed data (fire alarm or water-level data) must arrive successfully to the data repository. The maximum acceptable delay for considered data is 1 second.

IEEE 802.15.4 offers the possibility of retransmitting a packet if the source node does not receive an acknowledgment from the destination node. In a network with a huge number of nodes (similar to a shipboard WSN), the number of retransmissions has an important impact on the global performance of the network, including the packet delivery ratio, the end-to-end delay, the energy consumption of nodes and the network load.

Fig. 6 shows the evolution of the packet delivery ratio of the network with respect to the maximum number of retransmissions for the tree and mesh topologies. For the tree topology, the packet delivery ratio increases with the number of permitted retransmissions. It reaches 100 % when the retransmissions number is equal or higher than 10. It can be concluded from this curve that a maximum number of 10 retransmissions is sufficient to have a maximum packet delivery for the considered network. Otherwise, for the mesh topology, the packet delivery ratio increases rapidly until the number of retransmissions becomes 10 and decreases slowly for higher values. This may be explained by the collisions that can cause the retransmissions of failed packets. Therefore, a maximum value of 10 retransmissions is an optimal value for the two topologies.

It can be noticed in this figure that the packet delivery ratio achieves 99% for 8 retransmissions, which is equal to the average packet delivery found in the network test (8 retransmissions in the XMesh protocol). It is also seen in the figure that the packet delivery ratio is slightly higher for the tree topology. The particular ship environment makes this advantage of the tree topology.

Fig. 7 shows the variations of the average end-to-end delay with respect to the maximum number of retransmissions for the tree and mesh topology of ZigBee network. End-to-end delay is defined as the total delay between creation and reception of an application packet. This figure shows that the average delay increases when the maximum number of retransmissions increases. For the tree topology, the delay increases rapidly for a maximum retransmissions number lower than 10.

For larger values of the maximum number of retransmissions, its variations become small. This result is coherent with the packet delivery and confirms that 10 retransmissions are sufficient to have a reliable tree-topology network. The value of delay achieved is 0.1 second which is acceptable for the shipboard monitoring system that supports a maximum delay of 1 second. Otherwise, the delay keeps increasing in the case of mesh topology. It is slightly higher than the delay of tree topology. This is basically due to the differences in the routing techniques and the size of routing tables in the mesh topology where the route discovery procedure induces additional delays.

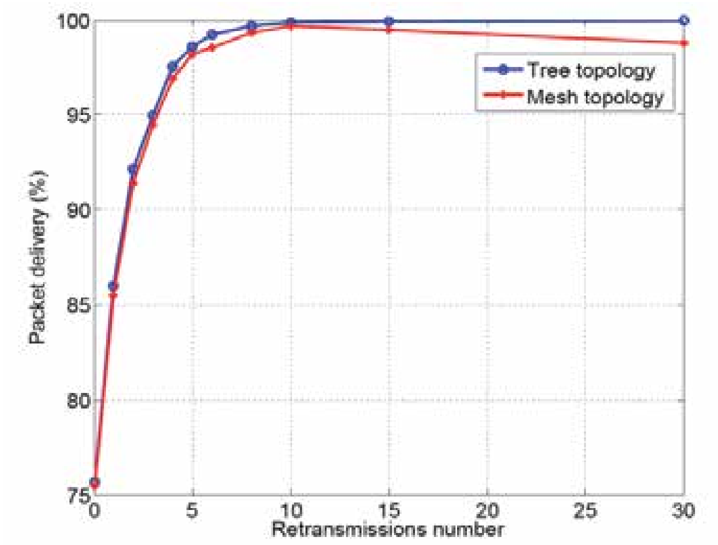


Figure 6. Packet delivery ratio versus the number of retransmissions

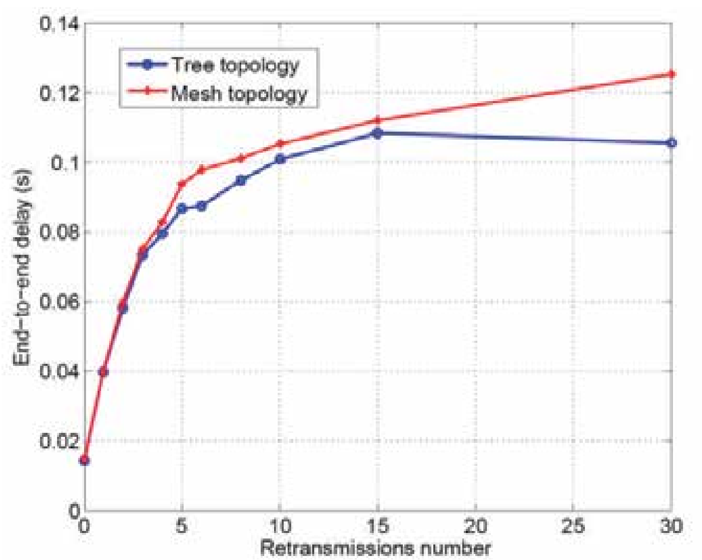


Figure 7. End-to-end delay versus the number of retransmissions

7. Conclusion

In this chapter, the application of wireless technologies to the shipboard monitoring system has been studied. A measurement campaign has been carried out on board a ferry to determine path loss models. An IEEE 802.15.4 compliant WSN has been tested successfully on board the same ferry. Based on the measurement results and the particularities of the environment, a hierarchical zone-based architecture has been proposed for a large shipboard WSN. The performance of this architecture has been evaluated using ZigBee standard. In order to obtain a reliable and representative simulation, the path loss models obtained from the measurement campaign have been integrated into the simulator. The obtained delay and packet delivery ratio meet the difficult requirements of the shipboard monitoring system. These results have also shown that ZigBee may be an appropriate technology for the proposed architecture.

In spite the successful tests in verifying the WSN's feasibility onboard ships, the introduction of wireless solutions in the shipboard monitoring system is not likely to happen quickly. Special attention must be given to the development of shipboard sensor nodes: this equipment must resist against hostile environmental conditions in the engine rooms such as temperature, humidity, vibration, etc. Additionally, several steps including testing, regulation and standardization will be necessary before deployment.

Author details

Hussein Kdouh, Gheorghe Zaharia, Christian Brousseau,
Hanna Farhat, Guy Grunfelder and Ghaïs El Zein
Institut d'Electronique et de Télécommunications de Rennes, Rennes, France

Acknowledgement

This work is a part of SAPHIR project which is supported by 'Région Bretagne' and 'Pôle Mer Bretagne'. The authors thank Marinelec Technologies and Compagnie Océane for the opportunity to conduct the measurement campaigns.

8. References

Arora, A.; Dutta, P.; Bapat, S.; Kulathumani, V.; Zhang, H.; Naik, V.; Mittal, V.; Cao, H.; Demirbas, M.; Gouda, M.; Choi, Y.; Herman, T.; Kulkarni, S.; Arumugam, U.; Nesterenko, M.; Vora, A. & Miyashita, M. (2004). A line in the sand: a wireless sensor network for target detection, classification, and tracking. *Computer Networks*, Vol. 46, No. 5, pp. 605 – 634, ISSN 13891286

Barrenetxea, G.; Ingelrest, F.; Schaefer, G.; Vetterli, M.; Couach, O. & Parlange, M. (2008). Sensorscope: Out-of-the-box environmental monitoring. *Proceedings of the IEEE 7th International Conference on Information Processing in Sensor Networks (IPSN '08)*, Washington U. St. Louis, Missouri, USA, April 22-24, 2008

Barrenetxea, G.; Ingelrest, F.; Schaefer, G. & Vetterli, M. (2008). The hitchhiker's guide to successful wireless sensor network deployments. *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys 2008),* Raleigh, North Carolina, November 5-7, 2008

Brown, W.M.; Watson, M.A. & Plonski, F.E. (2003). Shipboard wireless site selection techniques. *Proceedings of the 13th Ship Control Systems Symposium (SCSS)*, Orlando, Florida, USA, April 7-9, 2003

Durgin, G.; Rappaport, T.S. & Xu, X. (1998). Measurements and models for radio path loss and penetration loss in and around homes and trees at 5.85 GHz. *IEEE Transactions on Communications*, Vol. 46, No. 11, pp. 1484 - 1496

Estes, D.R.J.; Welch, T.B.; Sarkady, A.A. & Withesel, H. (2001). Shipboard Radio Frequency Propagation Measurements for Wireless Networks. *Proceedings of the IEEE Military Communications Conference (MILCOM), IEEE Communications for Network-Centric Operations: Creating the Information Force,* Tysons Corner McLean, Virginia, USA, October 28-31, 2001

Gao, T.; Greenspan, D.; Welsh, M.; Juang, R. R. & Alm, A. (2005). Vital Signs Monitoring and Patient Tracking Over a Wireless Network. *Proceedings of the 27th IEEE Annual International Conference on the Engineering in Medicine and Biology Society (EMBS)*, Shanghai, China, September 1-4, 2005

Kang, H.J.; Lee, D.; Shin, J.G. & Park, B.J. (2011). Location tracking of moving crew members for effective damage control in an emergency. *Defence Science Journal*, Vol. 61, No. 1, pp. 57–61, ISSN 0011748X

Kdouh, H.; Zaharia, G.; Brousseau, C.; Grunfelder, G. & El Zein, G. (2011). ZigBee-based sensor network for shipboard environments. *Proceedings of the 10th IEEE International Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, June 28 - July 02, 2011

Kdouh, H.; Brousseau, C.; Zaharia, G.; Grunfelder, G. & El Zein, G. (2011). Applying ubiquitous wireless technologies for shipboard monitoring systems. *Proceedings of the 14th IEEE International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Brest, France, October 03-07, 2011

Kdouh, H.; Brousseau, C.; Zaharia, G.; Grunfelder, G. & El Zein, G. (2011). Measurements and path loss models for shipboard environments at 2.4 GHz. *Proceedings of the 41th European Microwave Conference (EuMC)*, Manchester, UK, October 09-14, 2011

Kdouh, H.; Farahat, H.; Zaharia, G.; Brousseau, C.; Grunfelder, G. & El Zein, G. (2012). Performance analysis of a hierarchical shipboard wireless sensor network. Submitted to *23rd Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sydney, Australia, September 09-12, 2012

Krishnamurth, L.; Adler, R.; Buonadonna, P.; Chhabra, J.; Flanigan, M.; Kushalnagar, N.; Nachman, L. & Yarvis, M. (2005). Design and Deployment of Industrial Sensor Networks: Experiences from a Semiconductor Plant and the North Sea. *Proceedings of the 3rd international conference on Embedded networked sensor systems*, San Diego, November 2-4, 2005

Langendoen, K.; Baggio, A. & Visser, O. (2006). Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture. *Proceedings of the 20th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2006)*, Isle of Rhodes, Greece, April 25-29, 2006

Li, H.W; Zhou, Z. & Ou, J.P. (2003). Offshore Platform Health Monitoring Applications of Wireless Sensor. *Proceedings of the 21st International Modal Analysis Conference (IMAC XXI)*, Kissimmee, Florida, February 3-6, 2003

Lynch, J.P. & Loh, K.J. (2006). A summary review of wireless sensors and sensor networks for structural health monitoring. *The Shock and Vibration Digest*, Vol. 38, No. 2, pp.91-128, ISSN 0583-1024

Memsic Technology (2007). MicaZ motes datasheet. Retrieved from http://www.memsic.com/support/documentation/wireless-sensor-networks/category/7-datasheets.html

Memsic Technology (2007). Xmesh user's manual. Retrieved from http://www.memsic.com/support/documentation/wireless-sensor-networks/category/6-user- manuals.html

Mokole, E.L.; Parent, M.; Samaddar, S.N.; Tomas, E. & Gold, B. T. (2000). Radio-frequency Propagation Measurements in Confined Ship Spaces Aboard the ex-USS Shadwell. *Report NumberA987183*, Naval Research Laboratory

OPNET Technologies. (n.d.) available from http://www.opnet.com

Ou, J-P. & Li, H-W. (2003). Wireless Sensors Information Fusion for Structural Health Monitoring. *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications at the SPIE Defense and Security Symposium*, Orlando (Kissimmee), Florida, USA, April 23-25, 2003

Padhy, P.; Martinez, K.; Riddoch, A.; Ong, H.L.R. & Hart, J.K. (2005). Glacial environment monitoring using sensor networks. *Proceedings of the Workshop on Real-World Wireless Sensor Networks (REALWSN 2005)*, Stockholm, Sweden, June 20-21, 2005

Paek, J.; Chintalapudi, K.; Govindan, R.; Caffrey, J. & Masri, S. (2005). A wireless sensor network for structural health monitoring: Performance and experience. *Proceedings of the 2nd IEEE Workshop on Embedded Networked Sensors (EmNets 2005)*, Sydney, Australia, May 30-31, 2005

Paik, B-G.; Cho, S-R.; Park, B-J.; Lee, D.; Bae, B-D. & Yun, J-H. (2009). Characteristics of wireless sensor network for full-scale ship application. *Journal of Marine Science and Technology*, Vol. 14, No. 1, pp. 115–126, ISSN 1023-2796

Park, B-J.; Paik, B-G.; Cho, S-R.; Lee, D.; Kang, H. & Choi, J. (2008). Application of Ubiquitous Technology to Ship Environment. Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Dublin, Ireland, July 21-25, 2008

Perkins, C.E. & Royer, E.M. (1999). Ad-hoc on-demand distance vector routing. Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, USA, February1999

Pilsak, T.; Ter Haseborg, J.L. & Hanneken, H. (2009). WLAN propagation on the bridge of vessels under consideration of materials properties. Proceedings of the IEEE *International Symposium on Electromagnetic Compatibility (EMC Europe)*, Athens, Greece, June 11-12, 2009

Ploeger, R.; Newton , W.W.; Rabiner, A. & Lally, R. (2003). Wireless e-diagnostics reduces workload and improves shipboard quality of life. *White Paper, Oceana Sensor Technologies*.

Polastre, J.P.; Szewczyk, R.; Mainwaring, A.; Culler, D. & Anderson, J. (2004). Analysis of wireless sensor networks for habitat monitoring, In: *Wireless Sensor Networks*, Raghavendra, C.S.; Sivalingam, K.M. & Znati, T., pp. 399-423, Kluwer Academic Publishers, ISBN 1402078838

Rappaport, T.S. (2002). *Wireless Communications: Principles and Practice*, Prentice Hal, ISSBN 0130422320

Schwartz, G. (2002). Reliability and Survivability in the Reduced Ship's Crew by Virtual Presence System. *Proceedings of the IEEE International Conference on Dependable Systems and Networks (DSN'02)*, Bethesda, Maryland, USA, June 23-26, 2002

Selavo, L.; Wood, A.; Cao, Q.; Sookoor, T.; Liu, H.; Srinivasan, A.; Wu, Y.; Kang, W.; Stankovic, J.; Young, D. & Porter, J. (2007). Luster: wireless sensor network for environmental research. *Proceedings of the 5th ACM Conf. Embedded Networked Sensor Systems (SenSys 2007)*, Sydney, Australia, November 04-09, 2007

Stoianov, I.; Nachman, L.; Madden, S. & Tokmouline, T. (2007). Pipenet - a wireless sensor network for pipeline monitoring. *Proceedings of the 6th ACM International Conference on Information Processing in Sensor Networks (IPSN '07)*, MIT Cambridge, Massachusetts, USA, April 25-27, 2007

Takahashi, D. (2004). Dusting for Data. *TheMercuryNews.com*. Retrieved from http://www.mercurynews.com/mld/mercurynews/business/technology/9770734.htm

Tateson, J.; Roadknight, C.; Gonzalez, A.; Fitz, S.; Boyd, N.; Vincent, C. & Marshall, I. (2005). Real world issues in deploying a wireless sensor network for oceanography. *Proceedings of Workshop on Real-World Wireless Sensor Networks (REALWSN 2005)*, Stockholm, Sweden, June 20-21, 2005

Werner-Allen, G.; Lorincz, K.; Johnson, J.; Lees, J. & Welsh, M. (2006). Fidelity and yield in a volcano monitoring sensor network. *Proceedings of the 7th Symposium on Operating Systems Design and Implementation (OSDI '06)*, Seattle, Washington, USA, November 6–8, 2006

Yick, J.; Mukkerjee, B. & Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks*, Vol. 52, No.12, pp. 2292–2330, ISSN 13891286

ZigBee Alliance. (n.d) available from http://www.zigbee.org

Development and Implementation of Wireless Sensor Network for the Electricity Substation Monitoring

Paulo Sausen, Airam Sausen, Fabiano Salvadori, Renê Emmel Júnior
and Mauricio de Campos

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50489

1. Introduction

The requirements for process instrumentation, combined with the advances in wireless communications and electronics allowed the design of wireless sensor networks (WSNs). The technology applied to these sensors and communication networks has enabled the evolution of these systems has been called a smart sensor networks. In this case, not only collect sensor data, but also perform local processing, and may also act in the system, and subsequently, if necessary perform transmission. These smart sensor networks allows a more effective monitoring system, fault detection, and others, thus improving the reliability and system maintenance [2, 3, 6].

Among the challenges of design, development and installation ofsmart sensor networks, we can highlight environments where electromagnetic interference can reduce your performance or make it inoperable. In such cases, hybrid networks, that combine wireless systems with wired structures may be more appropriate [8]. These hybrid structures also allow better power management of these networks, since in some the cases the sensor node can be installed in places with difficult or no access. In these cases, the physical connection can even be used as a redundancy of the communication system.

In cases where the substation has been installed and is running without the provision of a monitoring system, it is not possible with commercial solutions available to monitor the system without performing some kind of reform of infrastructure. In this particular case, the structure reforms to allow passage of communication cables bring numerous disorders. Currently these power substations do not have any type of on-line monitoring. In cases where there a fault, after some consumers inform the concessionaire, a team must move to the place to search for the substation where the fault occurred.

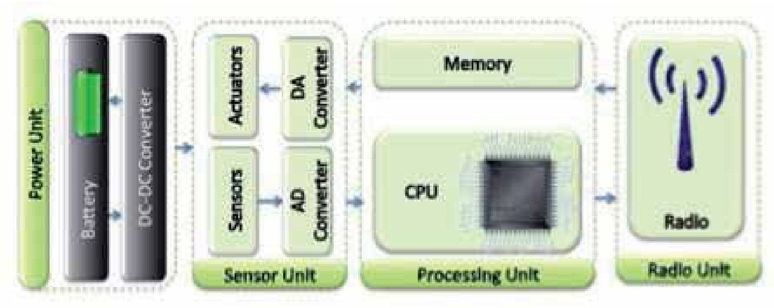


Figure 1. Sensor Node Basic Structure

The objective of this chapter is to develop a integrating system considering a set of smart sensors and communication systems for use in a underground power distribution substation. The underground power distribution substation chosen, owned to the network grid system in the city of Porto Alegre in Brazil. The depth of these substations is 4-5 meters, under layers of asphalt and concrete. Therefore another challenge of this work was to establish the communication with the system installed in the interior of substation and the outside, since this is not possible through radio systems and are not available physical bus installed for this purpose.

2. Wireless sensor networks

A Wireless Sensor Network can contain hundreds or thousands of small autonomous elements called sensor nodes, and each sensor can feature a large variety of sensors (e.g., temperature, speed, acoustic, seismic). In many cases, nodes are randomly spread over remote areas, making it difficult to perform any maintenance to the nodes. Hence, a node remains live while it has enough battery capacity for its normal operation, and the network lifetime strongly depends on the remaining capacity of the nodes in the network.A sensor node has a few basic components (see Figure 1) as follows [1, 7]:

- *Power Unit*, usually a battery, which acts as the power source for all node's components;
- *Sensor Unit* that contains a group of sensors and actuators;
- *Processing Unit* which includes a microprocessor or a micro-controller;
- *Communication Unit* which consists of a short range radio for wireless communication.

3. Network grid underground distribution system

The underground distribution systems represent an attractive alternative for applications in distribution systems in large urban centers, which are characterized by large concentrations of load and require high levels of quality, continuity and reliability of electricity supply.

There are two most common forms of connecting underground distribution systems, the radial system or the network grid system. The network system, is a low voltage distribution system having a set of transformers which the secondary side are connected in parallel,

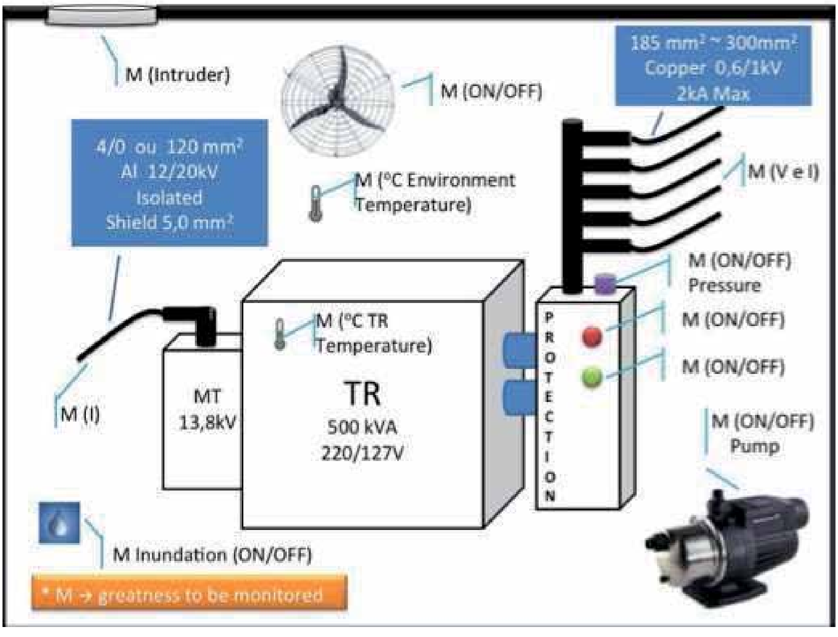


Figure 2. The Greatness Monitored by the Developed Systems

supplying the load. This topology allows the supply of electricity is maintained even if one or more transformers to leave the provided service, as long as the total power of the remaining transformers is equal to or higher than load demands, furthermore allows to improve the voltage characteristic of secondary.

The network grid system is installed at the central region of Porto Alegre is fed with a primary voltage of 13.8 kV and secondary voltages of 127/220V, being composed of a 500 kVA transformers, submersible, hosted in subterranean chambers, which are accommodated under of the central city streets. The biggest risks in this type of system are inundation, overheating, faults in the protection system, alterations in the pressure of the protection system. In the Figure 2 can be observed the greatness monitored by the developed system.

4. Developed architecture for monitoring system

The system developed (see Figure 3) is based on the concept of intelligent sensors. The Smart Sensors Modules (SSM's) can take a reading of up to four greatness, two analog and two digital, communicating through a wireless network and/or a physical network [4, 5].

A second module is designed for use with the acquisition of the greatness, this has a quickly dynamics and need to read more than four channels such as: voltages and currents in the three fase system in the secondary of the transformer. This device is called as remote data acquisition unit (RDAU). The data of these two systems are concentrated by the developed Gateway. The Gateway establishes communication with the exterior. As stated before, this communication is not possible through a radio link or a wired conventional structure since the characteristics of the substation does not allow the deployment of these systems. Therefore

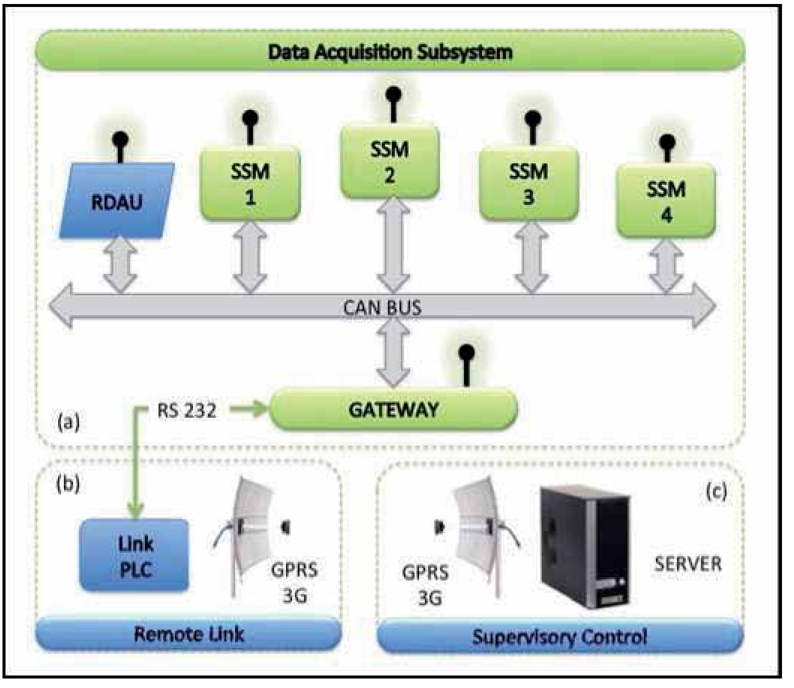


Figure 3. The Monitoring System Developed.

it has been used a PLC (Power Line Communication), allowing extraction of data from that environment (i.e., Underground Substation).

In the outside area the received data is retransmitted by a system GPRS/3G to a server. The monitoring system presents with the following subsystems (see Figure 3):

- Data Acquisition Subsystem;
- Remote Link; and
- Control Subsystem.

5. Intelligent sensor module - SSM’s

The SSM’s devices (See Figure 4) are capable of performing functions of sensoring, processing and transmitting/receiving data. Its architecture (See Figure 5) consists of a power subsystem, a sensor subsystem and communication subsystem.

The sensor subsystem and communication subsystem are managed by a PIC18F2580 microcontroller. This was chosen because of project requirements and also have by the same have be a built hardware dedicated to the CAN bus communication. In addition, adds support to various peripherals, such as analog-digital converter (A / D) 10-bit, four timers, USART serial interface (UNI-sectional sSynchronous Asynchronous Receiver Trans-Mitter), among others. The power subsystem is responsible for maintaining the supplied for the SSM, being the primary source of energy from the CAN bus and / or battery pack. When needed, the

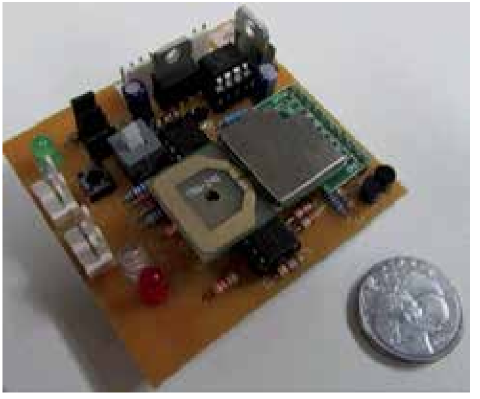


Figure 4. Intelligent Sensor Module - SSM.

CAN bus also feeds the recharging batteries system. This system consists in a battery pack with a capacity of 900 mAh and 7.2 V.

The SSM is equipped with four sensor inputs, 2 digital and 2 analog. The analog inputs are pre-conditioned to receive signals in the range 0-5 V or 4-20 mA, depending on the characteristics of the sensor connected. If the sensor connected to the SSM need power supply, it is provided together with the signal connector.

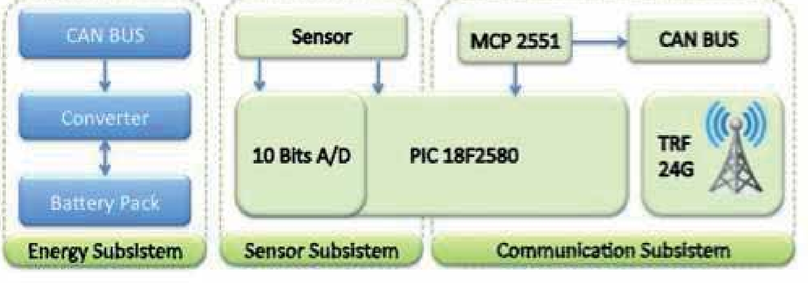


Figure 5. SSM Architecture.

The SSM using the communication subsystem to send / receive data in two different ways: via wireless network or via the physical network. The physical network is primarily intended for the redundancy, the wireless network being the primary means of exchanging information. As a communication device by radio frequency module was used TRF-24G, which employs the transceiver nRF2401A. This device performs the modulation GFSK (Gaussian Frequency Shift Keying) in data transmission at a rate of a maximum 1Mbps. It features integrated antenna and the transmission power can be set in the range -20 to 0 dBm, enabling a range of 250 m (without obstacles).

The physical bus meets the norm ISO11898-2 international standard intended for the CAN communication [3]. It specifies patterns relating to the physical layer of the CAN protocol, one being the use of a transceiver device that makes the interface between the sensor and CAN bus node, making certain electrical conditions provided in the standard are met.

Figure 6. Remote Data Acquisition Unit - RDAU.

Among these conditions include the protection against short circuits, voltage levels, among others. Therefore, the SSM's are connected to the bus via the CAN transceiver MCP2551 manufactured by Microchip Technology. A prototype of SSM was developed to experimental validation. Every SSM has an address assigned by the Gateway during installation the network, organizing themselves independently (plug and play).

6. Remote Data Acquisition Unit - RDAU

Differently from ISM topology the RDAU (see Figure 6) is a static projected acquisition unit for monitored system, with three voltage and three currents acquisition, transmission to RS, besides incorporate four digital inputs.

Voltage and current transducers

For voltage and current signals acquisition the transducers LV20-P and LA25-NP (LEMTM) are used respectively (hall- effect sensor). The voltage and current RMS phase values in the secondary of the measurement transformer are, respectively, 65V and 5A. Considering 50% overvoltage and 150% overcurrent, the system is able to measure until 100V (voltage) and 12,5A (current). To prevent aliasing problems, it was added a anti-aliasing, Butterworth filter is the 2nd order low-pass.

The Digital Signal Processor - DSP

The TMS320F2812 DSP uses Harvard architecture, with: 64kB of program memory, 64kB of data memory, 18kB of RAM memory and 1MB external interface memory. The operate frequency is 150MHz (6.67ns/instruction).

DSP software

The TMS320F2812 software was developed in C++ language, using Code Composer PlatinumTM platform (Texas Instruments TM). The program was divided in sub-routines, facilitating the agreement, maintenance and posterior update.

Main routine

This routine configures peripherals like GPIO, interruptions and external components configuration, as transceiver. The pins had been configured as inputs and outputs, except for the SPI pins and external interruption pins. A/D Control Routine The A/D converter is started by Adc.h library. The AD conversion routines had been configured with auto conversion of six samples in simultaneous sampling cascaded sequencer mode. It becomes simultaneous acquisition of one voltage phase and one current phase, 80ns dephasing phase/phase. The clock frequency is 25MHz. Communication Routines The communication can be carried out by two modules: asynchronous (RS-232), using SCI port; or, synchronous (wireless), using SPI port. At radio transmission, are sent 68 samples/cycle, while in serial transmission RS 232 are sent 10 samples/cycle. In asynchronous communication (serial RS-232), the transmission routine was configured with 1 stop bit, none parity, 8 data bits and speed transmission of 115200bps. The samples are grouped in a package forming a synchronism protocol, indicating for the receiver the beginning of the package and permitting to verify if the received data are correct. The package is formed by 4 heading words, 1 checksum data word, 1 checksum heading word and 60 data words. In the synchronous communication routine (to transceiver), SPI port is configured as a master, sending 16 data bits, with 1MHz transmission tax. This routine also defines the clock polarity, data transitions are during the rising edge, but delayed by half clock cycle, in agreement with the requirements of the transceiver. Besides configuring the reception interruption. The transmission routine is qualified each two conversions carried out by A/D converter. In this routine the data package are grouped to be sent to transceiver (transmitting mode). The first byte is preamble, the second address, followed for 14 bytes of data and finishing with CRC16, that transceiver (receiver) uses to verify if the received data are correct. The reception interruption is generated each time that the receive buffer (FIFO) will have 14 data bytes and the DR1 will be enabled.

Transceiver module

In the same way that ISM, the RDAU uses TRF-2.4GTM transceiver module for radio communication. The communication between the microprocessed unit and the transceiver is carried out through synchronous serial interface. The bits are sent to radio or received from a defined tax of clock. The TRW-24G configuration was carried out separately, one as receiving and other as transmitting. The receiver was configured as ShockBurst operation mode, with 24bits address, one reception canal, 14 data words and 16bits CRC. The transmitter was configured as direct mode, because in the ShockBurst mode the transceiver must be disabled each final data package, besides there is 50ns delay to each time that this is active again. In the direct way, always active, the data are only sent if there is a clock together the data. After the configurations the main routine begins an infinite loop, waits the interruptions and to each 2 A/D conversion begins the sending package routine.

7. Gateway

It was also developed a Gateway which is responsible for interconnecting all sensors (SSM and RDAU) and the transmission system PLC. The essential difference of the Gateway to the SSM is that there are an additional RS232 serial communications port used to perform the interconnection with the PLC. The physical aspect of the Gateway is shown in Figure 7.

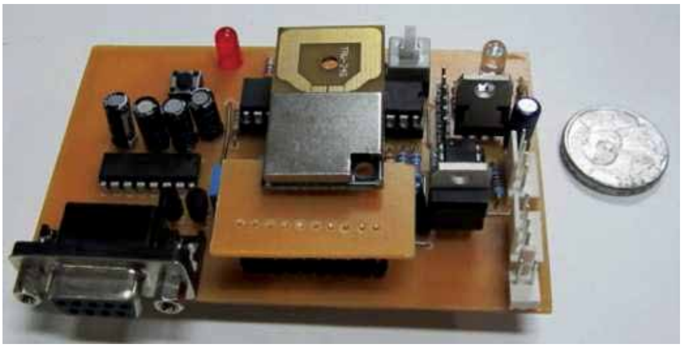


Figure 7. Gateway.

The exchange of information between the Gateway and SSM's is done over the MODBUS communication protocol. This protocol is characterized essentially by being the master-slave. Defines a structure of communication messages used to transfer data between analog and discrete devices with microprocessor detection and information transmission errors. The MODBUS protocol is located at 7th level of the OSI Reference Model (Open Systems Interconnection), which corresponds to the application layer that provides communication such as "client / server" between devices connected to different types of buses or network topologies [4]. The MODBUS also allows easy integration with SCADA systems although not the main focus of this work.

The management and addressing SSM's is performed by the Gateway, which in turn, updates and constantly checks the presence of new SSM's that are connected to the bus.

8. PLC modem

The PLC system installed in the the central area of Porto Alegre city, in low voltage cabling, underground network, consists of a pair of transmitter / receiver PLC, developed from a PLC MODEM PL-3120 manufacturing ECHELON. MODEM connected to the PL-3120 is a microcontroller whose functions are:

- Transmitter / Receiver PLC installed on the transformer;
- Data collection and ambient temperature of the transformer housing;
- Generation of data packet to send to the MODEM PL-3120 through the serial interface (UART); and
- Management control messages sent via the mains, supplied by MODEM PL-3120.

Transmitter / Receiver PLC installed in the former tender:

- Receipt of data packets sent via the mains, supplied by MODEM PL-3120;
- Check validity of data received;
- Configuring the Modem GSM / GPRS;
- Generation of data packet to send to the GSM MODEM / GPRS through UART serial interface; and

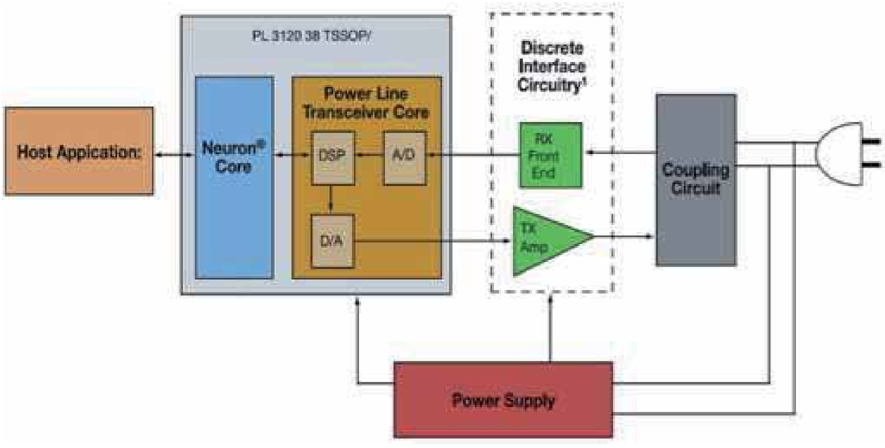


Figure 8. PLC Node based in PL-3120.

- Management control messages sent over the cellular network, delivered by the GSM MODEM / GPRS.

The PLC MODEM PL-3120 incorporates a NEURON processor, 4Kbytes of application memory and 2Kbytes RAM. The processor performs the NEURON routines protocol interconnecting the nodes of a network PLC, ISI - Interoperable Self Installation, and communication protocols, with an option to activate or not the protocol CENELEC. All these protocols are proprietary and has been recorded in ROM on the device. Figure 8 presents a block diagram of the constituents with a PLC node based on the PL-3120.

The MODEM PL-3120 can operate in bands A and C defined in CENELEC standards, which are selected from the crystal used to trigger the MODEM. The selection of the band CENELEC also defines the rate of data transmission on the electrical system. By selecting the band, the communication will occur at a rate of 3.6Kbps.

As shown in block diagram in Figure 8, there is the need for integration of an interface between the PL-3120 and the circuit will make the coupling of the modulated carrier to the grid. The interface circuit is composed mainly of an amplifier that can apply a signal to the power supply in one of the frequencies of operation of the PL-3120, with up to 1A peak-to-peak. Figure 9 shows the circuit diagram of the amplifier output, which forms part of the interface circuit. It is a discrete transistor circuit in a modified push-pull configuration.

The Figure 10 presents an analysis of frequency response of the power amplifier of the PLC transceiver. It can be seen a response practically flat in the frequency range of 1kHz to 20kHz. In the frequency range corresponding to the band The CENELEC standart there is a peak in the curve of the amplifier gain, the maximum occurs at a frequency of 100kHz, falling abruptly after this frequency.

The tested system was installed in the network system of CEEE-D (Companhia Estadual de Distribuição de Energia Elétrica) in the metropolitan area of Porto Alegre city (See Figure 11). The monitoring system (RDAU and SSMs) was installed in the northeast network grid system (RNE), the transformer station T-103-7A (code CEEE-D), which is the feeder 2RNE as supplier

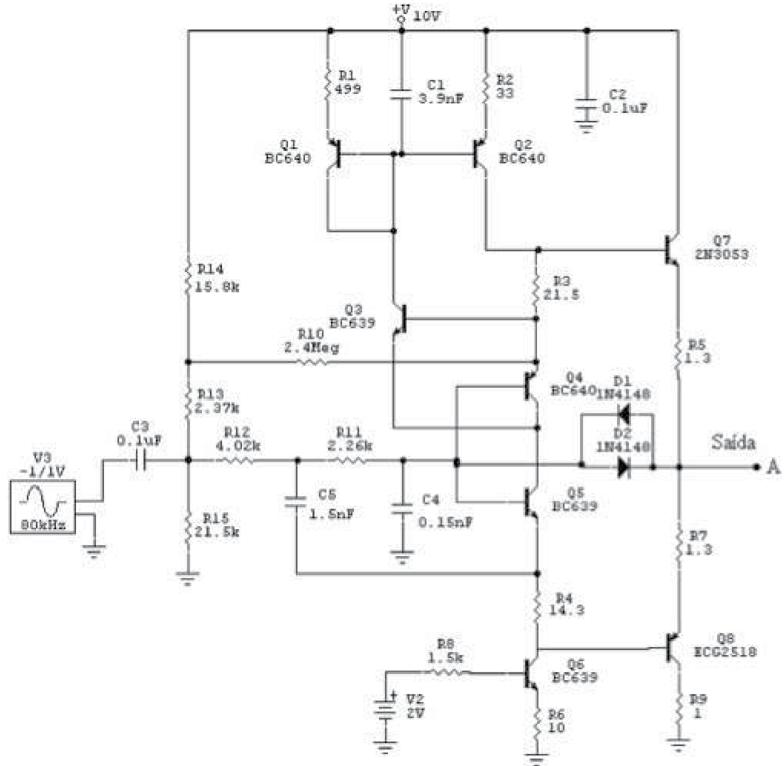


Figure 9. TX Amplifier.

of energy. The gateway developed manages the receipt of the data system and is connected to the PLC signal transmitter, the output low voltage transformer. The approximate distance between the transmitter and receiver is about 250 meters, since there is no direct path and the cables contour the square

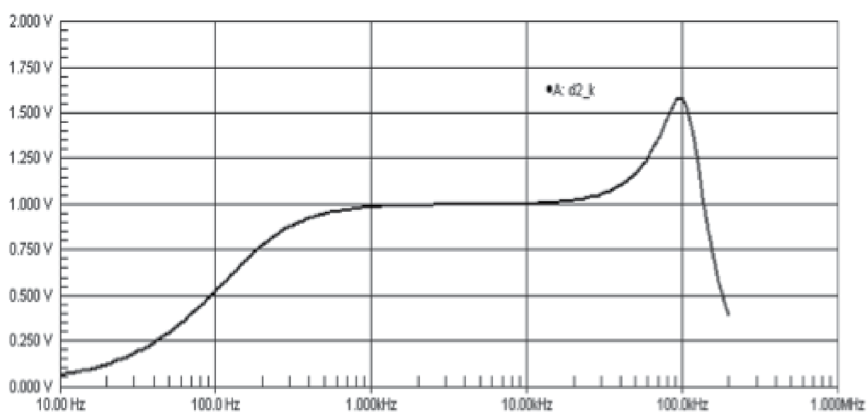


Figure 10. Frequency response of the TX amplifier.

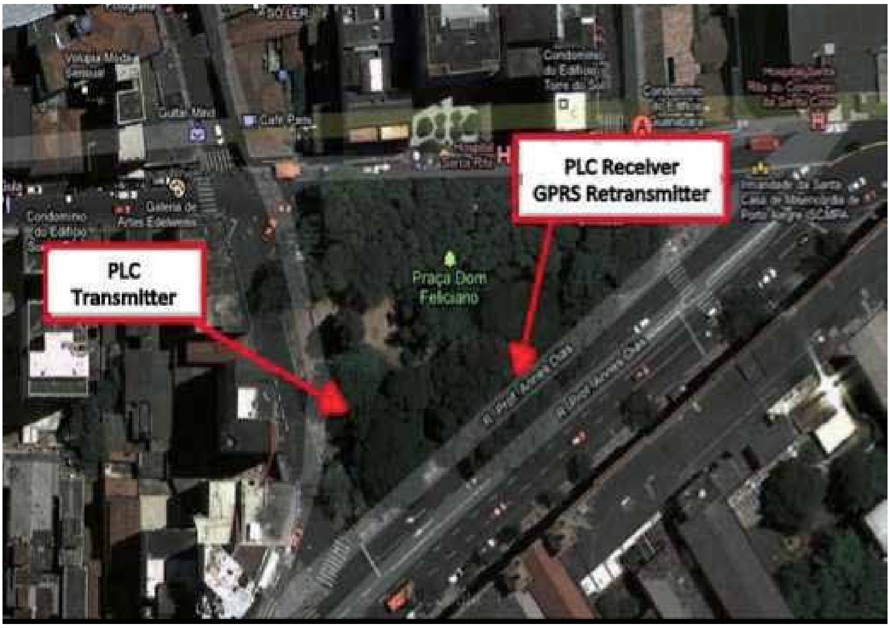


Figure 11. PLC Receptor and Transmitter localization.

Due to the robustness provided mainly by the adoption of structure hybrid in Smart Sensors Modules and Remote Data Acquisition Unit, there are no packet losses in communication between them, since the most critical data such as voltage and current traveling through the system CAN when they are not received appropriately by the tranceivers. The battery system of the sensor nodes acts as backup in cases where redundancy happens. In these cases, the worst possible condition, occurs when the sensor node is continuously processing and transmitting data and energy consumption reaches 57mA peak. Just as a tested set of batteries used in these extreme conditions that could be evaluated for its durability. The figure below shows these results.

The interval between samples is 1 second and therefore can be estimated operating time critical state in 4 hours and 15 minutes on average. After collecting the data on the server, they can be viewed in real time on supervisory controller.

The Supervisory Controller (SC) Subsystem was developed in order to emphasize data presentation versatility. It runs on standard PC architecture originally using a web browser and mobile devices with Android and a IOS respectively . The Application was developed to support on-line data presentation through reading of the received data files. Users may be able to explore, navigate and visualize all the data they are interested in. Since the main function of the SC subsystem is receive and store the data derived from the ISM?s and the RDAU?s, it can process and present them to operators through: i) SCADA (proprietors systems); or, ii) Man Machine Interface (HMI) specially designed for this purpose. Several possible operation scenarios can be considered for the SC subsystem. First, lets consider a scenario where the monitored data are distant from the SC and a wired infrastructure is available (Ethernet TCP/IP). Then, the acquired data for the RDAU and the ISM?s will be transmitted through the wired network communicating through the Remote Server (RS) . Another situation, where

Figure 12. Web based real timer viewer for the supervisory system.

the monitored data is distant but wired network does not exist. Then, the SR-SC connection can be made over Wi-Fi. In a third scenario, where the data monitored is close to the SC the transmission can be directly made - gateway to SC. The connection between gateway - SC is accomplished through serial communication, using either USB or RS-232 ports of the computer.

Figure 13. ANDROID real timer viewer for the supervisory system.

Figure 12 show the Human Machine Interface (HMI) is used in the SC was developed in PHP Language (Hypertext Preprocessor). Initially it was developed for the voltage and currents monitoring; temperature monitoring; pressure monitoring; and fifteen digital inputs for determination of state operation system. The HMI was extended to the smartphone platform aiming at facilitating access information from any mobile phone. In this context two applications have been developed (see Figure 14 and Figure 13) for the two platforms most used: Android Platform and Platform IOS.

Figure 14. IOS real timer viewer for the supervisory system.

9. Conclusions

This chapter has presented a monitoring system that is applied to underground power electrical substation. The advances in wireless communication, microelectronics, digital electronics, and highly integrated electronics, in addition to the increasing need for more efficient controlled electric systems, make the development of monitoring and supervisory control tools the object of study of many researchers. The system presented in this chapter allows choosing the desired communication transmission mode: wired, wireless, or both. Performance results show that our system could well be applied for monitoring and fault detection in electrical underground network grid systems.

The main advantages of this system are numerous. Firstly it allows the automation and monitoring of a substation without the need for any change infrastructure or other civil works. Second, because it is characterized as a non-invasive system, any device has to be drilled (one should remember that all equipment in the substation, they have IP68 protection - against floods). Finally, the redundancy possible by the choice of hybrid architecture allows the system proposed has high fault tolerance in the monitoring. The disadvantages of the proposed system is the fact that even using the 3G or 4G infrastructure as a means of long distance communication. In order to reduce the possibility of flaws in this system, two different cell phone carriers are used redundantly.

Author details

Paulo Sergio Sausen and Airam Sausen
Master's Program in Mathematical Modeling, Regional University of Northwestern Rio Grande do Sul State (UNIJUÍ), Brazil

Mauricio de Campos
Industrial Automation and Control Group (GAIC), Regional University of Northwestern Rio Grande do Sul State (UNIJUÍ), Brazil

Fabiano Salvadori
Federal University of Paraíba (UFPB), Brazil

10. References

[1] Akyildiz, I. F., W.Su, Sankarasubramaniam, Y. & Cayirci, E. [2002]. Wireless sensor networks: a survey, *IEEE Transactions on Computer Networks* 38(7): 393–422.

[2] Bricker, S., Gonen, T. & Rubin, L. [2001]. Substation automation technologies and advantages, *IEEE Computer Applications in Power* 14(3): 31–37.

[3] Gungor, V. C., Sahin, D., Kocak, T., Ergüt, S., Buccella, C., Cecati, C. & Hancke, G. P. [2011]. Smart grid technologies: Communication technologies and standards., *IEEE Trans. Industrial Informatics* 7(4): 529–539.
URL: *http://dblp.uni-trier.de/db/journals/tii/tii7.html#GungorSKEBCH11*

[4] Lakshmikanth, A. & Morcos, M. M. [2001]. A power quality monitoring system: A case study in dsp-based solutions for power electronics, *IEEE Transactions on Instrumentation and Measurement* 50(3): 724–731.

[5] Salvadori, F., Campos, M., de Camargo, R. F., Gehrke, C., Rech, C., Sausen, P. S., Spohn, M. A. & Oliveira, A. [2007]. Monitoring and diagnosis in industrial systems using wireless sensor networks, *Proceedings of the IEEE International Symposium on Intelligent Signal Processing (WISP)*.

[6] Salvadori, F., de Campos, M., Oliveira, A. C., de Camargo, R. F., Gehrke, C., Rech, C., Spohn, M. A. & Sausen, P. S. [2009]. Monitoring in industrial systems using wireless sensor network with dynamic power management, *IEEE T. Instrumentation and Measurement* 58(9): 3104–3111.

[7] Sausen, P. S., Sousa, J. R. B., Spohn, M. A., Perkusich, A. & Lima, A. M. N. [2007]. Dynamic power management with scheduled switching modes in wireless sensor networks, pp. 1–8.

[8] Sharma, G.; Mazumdar, R. R. [2008]. A case for hybrid sensor networks., *IEEE/ACM Transactions on Networking* 16(5): 1121–1131.

Chapter 15

Wireless Sensor Networks to Improve Road Monitoring

M. Collotta, G. Pau, V.M. Salerno and G. Scatá

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48505

1. Introduction

In this chapter an approach to improve road monitoring through wireless sensor networks is described: several sensors may be distributed along a road in order to detect traffic flows, speed, and the continued occupation of the road. Each WSN cluster offers a monitoring system based on videotapes and, at the same time, produces information which can be processed to provide a clear interpretation of the road situation regardless of weather conditions [1, 2]. Therefore, it is necessary to integrate video monitoring information with parameters measured by other sensors, e.g. magnetometer or power sensors, for traffic detection [3–5]. Advantages about WSN usage in this area are several. Some of them have been already shown in literature. Others derive by the proposed approach.

- A WSN can monitor and evaluate roads automatically and continuously, with little human effort
- A WSN can work during night even with poor weather conditions, when there is fog or presence of dust (pollution, volcanic ash) in the air
- WSNs are low cost and low power [6]
- A WSN allows the integration of video monitoring with magnetic or power sensor. In this way, it is possible to obtain complete and integrated information (video-images and traffic volumes information) [7]
- WSNs allow dynamic changes to network topology based on real needs and reports coming from sensors located along the road.
- When needed, the number of cameras which control a specific area may increase to produce more detailed information. However, it may increase network workload that will be properly managed by the proposed approach.

The idea is to distribute many sensors along a road to detect traffic flows, speed and the continued roadway occupation. At first, the chapter will show most common wireless technologies used in WSNs in order to select the protocol to meet main system requirements.

A system for road traffic monitoring, operating in adverse environmental conditions, involves the coexistence of a large number of devices that, working together, ensure proper analysis of vehicular traffic situation. Our goal is to show a flexible architecture in which information of specific road sections are provided and, in case of queue or traffic jam, a mechanism for traffic detection is activated. In particular, this chapter introduces an algorithm that, based on traffic volume measured values through magnetic sensor located on the ground, dynamically enables or disables cameras according to the real need to monitor the given area. Moreover, in order to manage power consumption, the proposed architecture is based on an innovative approach, using IEEE 802.15.4/ZigBee standard protocol [8]. As described in next section, IEEE 802.15.4/ZigBee is able to satisfy main system requirements. The proposed algorithm dynamically varies active devices number based on loading conditions through a fuzzy logic controller. The role of fuzzy logic controller is useful in terms of network load and power consumption management. In order to satisfy several Quality of Service requirements (QoS) and real-time constraints, the network coordinator dynamically manages the data sampling period of cluster coordinator nodes. They give the new sampling period to its end-nodes representing local monitoring stations. The algorithm is characterized by predetermined "membership" functions used to process some network information (e.g. Deadline Miss, Power Consumption). The architecture proposed and the algorithms shown will be evaluated using TrueTime [35] and OmNet++ [36] simulators.

2. System requirements and wireless sensor network protocols

2.1. System requirements

In order to make a wireless sensor network suitable for road monitoring, some requirements have to be met. In the following we give a brief overview of requirements that drove the design of the proposed architecture and the choice of the wireless protocol to use.

- Predictability and system performance simulation: The system shall implement a WSN that allows to simulate its network environment and to determine in advance end-to-end performance of the monitored system: power consumption, end-to-end latency (min, max, average), jitter, throughput.
- Quality of Service provisioning: The WSN shall implement advanced QoS mechanisms and a clear policy to ensure guaranteed performance. QoS will be ensured also when nodes number increase.
- The network should cover harsh large area: The WSN should be able to work in a harsh and dynamic environments taking into account factors like high temperature, dust, vibrations, humidity, poor visibility, fog or heavy rain, etc.
- High density of nodes: Nodes density should be high.
- High communication reliability: The WSN shall provide high reliability in terms of communication services. Error message rate will be kept acceptable for the road monitoring application.
- Fault tolerance: The WSN shall prevent performance degradation in case of fault to any part of the system. It is mandatory to conduct a fault analysis of the monitored system in order to provide fault tolerance mechanisms. The WSN shall be self healing: i.e. the WSN detects communication errors and heals these errors by its own means.

- Mechanism to support dynamic environment: The working conditions in road monitoring systems are not static.
- Network ability to transmit video traffic flows: WSNs should be able to transmit video traffic flows using an appropriate video compression algorithm. The WSN must be able to manage QoS associated with this type of traffic flows.

2.2. Wireless sensor network protocols

Most important standard for communication in WSNs are IEEE 802.15.4, IEEE 802.15.4/a, IEEE 802.15.1 Bluetooth, 6LoWPan and WilressHART. Among these protocols IEEE 802.15.4/ZIgBee and IEEE 802.15.1 Bluetooth have been extensively explored. Even if they were born with the same purpose (to define a standard for small wireless networks) IEEE 802.15.4/ZigBee [8] and IEEE 802.15.1 Bluetooth [9] are characterized by several differences. In the following, a brief overview of these protocols.

- Bluetooth is an economical and secure standard (IEEE 802.15.1) to exchange information among devices through a short-range radio frequency. Bluetooth operates between 2.4 and 2.5 GHz (ISM) frequencies, using a FSK modulation with a data rate of 720Kbps. Simple topologies can be realized using Bluetooth devices. Bluetooth networks are called piconets and are characterized by eight elements: a master and seven slaves. The master node works as network coordinator and performs clocks synchronization. Slave nodes work as passive nodes, accepting conditions coming from the master node. Each device can use adaptive transmission in order to save batteries.
- IEEE 802.15.4 standard protocol has been defined for short-range, low cost, low speed and low power wireless communications. It is useful for WPANs characterized by low bit rate and devices powered by batteries which can not be frequently replaced, e.g. sensor nodes. Its main features are: coverage of about 50-100 meters; maximum data rate of 250 Kbps; low transmission power and lower power consumption; a network level to ensure routing procedures; use of ACK messages to perform data retransmission in case of absence of receipt or transmission error.

Bluetooth has been thought for audio applications by creating a Frequency Hopping Spread Spectrum scheme (FHSS), and a master/slave protocol. IEEE 802.15.4/ZigBee, instead, focuses the attention on sensors and controllers. It is based on short messages and a Direct Sequence Spread Spectrum system (DSSS). In this case, to decide the reference protocol, it is important to remember the central theme of this work: wireless sensor networks. Bluetooth is characterized by higher data rate and a massive device diffusion, which can be integrated in all the latest laptop models, mobile phones and PDAs. However, IEEE 802.15.4/ZigBee has been designed for low power consumption (batteries of these devices can last several years) and for a simple and fast communication among terminals. So, IEEE 802.15.4/ZigBee is the best candidate for sensor networks development and presents new and interesting perspectives in this field.

2.2.1. IEEE 802.15.4 ZigBee

The standard [8] defines physical layer (PHY) and datalink specifications (MAC) in order to ensure low data-rate wireless communications among devices requiring low power consumption. A LP-WPAN may include two different devices: FFD (Full Function Device) and RFD (Reduced Function Device). An FFD node can operate as network coordinator, as cluster coordinator or as simple communication terminal. An FFD (Coordinator and Router) can communicate with all devices, while an RFD (End Device) can communicate with an FFD only. Main features provided by Physical layer (PHY) are radio transceiver activation and

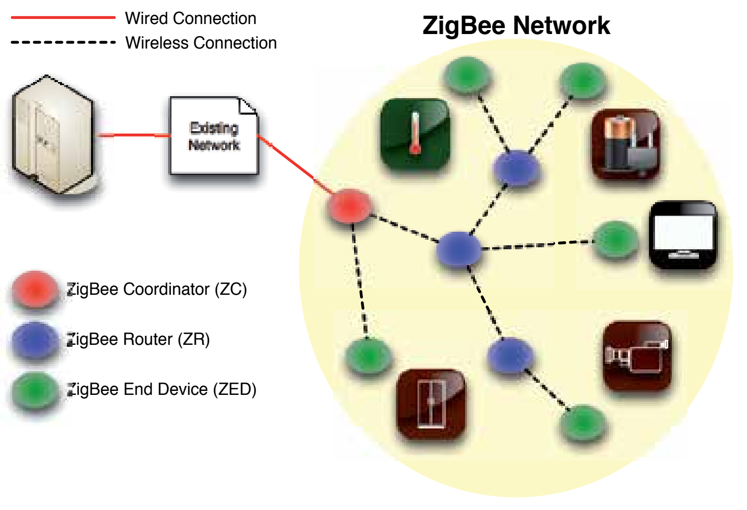


Figure 1. IEEE 802.15.4/ZigBee Network

deactivation, energy detection (ED), link quality indication (LQI), channel selection, channel availability estimation and the transmission and reception of packets through the physical medium (radio channel). The standard offers two possibilities for transmission/reception based on different frequencies. Both methods use DSSS (Direct Sequence Spread Spectrum) modulation technique. Transmission rate is 250 kbit/s (2.4 GHz), 40 kbit/s (915 MHz) and 20 kbit/s (868 MHz). Between 868 ÷ 868.3 MHz there is only one channel, while there are 10 channels between 902.0MHz and 928.0MHz. Finally, between 2.4GHz and 2.4835GHz there are more than 16 channels. The IEEE 802.15.4 supports the ability to dynamically select the channel using a scanning function that allows to search the beacon frame (synchronization) in a useful channels list. The main characteristic of the IEEE 802.15.4 PHY layer allows to satisfy power consumption requirements. In fact, it allows to perform other tasks, including receiver energy detection (ED). It is a feature used by the network layer for channel selection. MAC layer performs several tasks including: Mean access coordination; Packets creation and forwarding; Generation and address recognition; Packets sequence number control. It must also manage nodes detection process (Discovery). Time required to do this is about 30 ms, while other technologies like Bluetooth, need 5-6 s before they can fully use a device. Four frame types at MAC layer are possible: Data Frame; ACK Frame; MAC Command Frame; Beacon Frame. Data Frame consists of a maximum of 128 bytes numbered to ensure all packets routing. The "Frame Check Sequence" ensures that all packets are received without errors. This greatly improves transmission reliability in adverse conditions. Another important frame is the ACK frame. It provides confirmation that the sent packet has been correctly received. This solution ensures data consistency, but obviously increases latency. The MAC command frame provides a mechanism to monitor and configure client nodes. Finally, the beacon frame

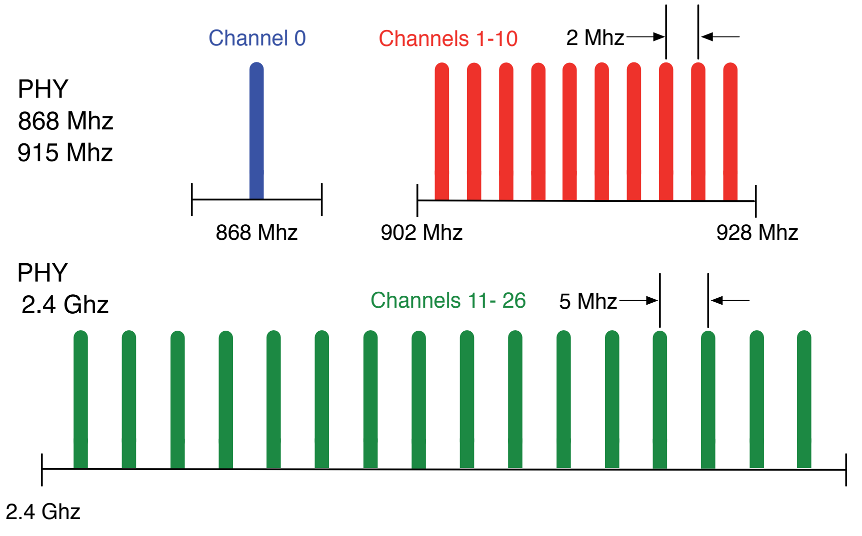


Figure 2. Physical layer

"wakes up" client devices, which are listening to their address and go into "sleep" mode if they do not receive it. Beacons (which are basically sync signals) are important for mesh and cluster tree networks to keep all nodes synchronized in order to save batteries. Since transmission medium (radio) is shared by all devices, it is necessary to use a transmission mechanism to avoid that two devices send packets simultaneously. There are two techniques: CSMA-CA and Beacon. Through CSMA-CA each device, before starting a transmission, performs a channel listening to understand if there is already another transmission. If so, the retransmission will be done later with a random delay. Through Beacon technique the coordinator sends a superframe (beacon mode) at regular time intervals (multiples of 15.38 ms, up to 252 s). Between a beacon and another one there are 16 time slots where the absence of collision is guaranteed. All devices contend for first 9 time slots, while remaining time

Figure 3. MAC frame structure

slots are assigned by the coordinator to a specific node and are called GTS (Guaranteed Time Slot). If a node has to transmit many information, the coordinator may give it more than one GTS. This structure guarantees dedicated bandwidth and low latency compared to the first technique. Furthermore, it also reduces battery consumption, because each device knows exactly when to transmit and it is sure there will not be collisions.

3. Video monitoring over wireless sensor networks

3.1. Introduction

Main requirement of a road monitoring scenario is the ability to guarantee Quality of Service for video monitoring traffic flows. However QoS management in a wireless sensor network is something difficult to ensure because of several factors:

- Sensors with limited power, energy, memory and processing capacity
- Nodes heterogeneity: each node should manage different tasks and different hardware
- Network topology may change over time due to nodes mobility, removal or addition of a node
- Environment conditions in which sensor work are often unpredictable

All these situations can cause "delays" and/or "packet loss rate" too high which damage predictability and reliability system requirements. It is therefore necessary to use a QoS management paradigm to ensure network flexibility, adaptability and scalability. Some of these problems have been already investigated in literature (par. 3.2) while others are unresolved yet. Finally, the last paragraph of this section shows video transmission over IEEE 802.15.4/ZigBee.

3.2. Road monitoring through WSNs in literature

Traffic congestion is a problem of many big city of the world since vehicles number increase. Research in Intelligent Transportation System (ITS) is often focused on optimization of urban traffic flows. ITS main goal is to improve safety, mobility and efficiency. In recent years researchers studied and analyzed wireless sensor networks, communication technologies and algorithms targeted for ITS applications. WSNs can be deployed and organized in a very short time and can measure sevaral parameters related to vehicular traffic. Furthermore, sensors nodes have reduced size, are often low power and can be positioned in various places along the road. A wireless sensor network is able to detect vehicular flow, speed and occupation with high spatial and temporal resolutions, providing a possible solution to traffic congestion. To improve road monitoring reducing, at the same time, waiting and travel time, techniques for efficient traffic management are necessary. A large number of methods and approaches have been proposed in literature to reach this goal. In [11] the effectiveness of a WSN to take real-time traffic information has been studied. A first analysis is performed using a single sensor to detect vehicles approaching to traffic lights. In a second analysis, two sensors are positioned on each traffic lane to calculate the queue length. Results show that the use of one sensor is not enough to obtain best performance. Instead the distance between two sensors, located in the same traffic lane [12], does not have a great influence on traffic flow. A sensors based Traffic Light Controller (TLC) is implemented thanks to the Green Light District Simulator [4] and it is compared with an existing real TLC that does not use sensors. Results show that the proposed sensor network based TLC, has good performance but it was never able to draw up the exact number of queued vehicles at the traffic lights. In fact, the use of not many sensors does not guarantee a full and complete management of traffic in a road intersection. A new decentralised TLC based on WSNs is proposed in [13]. Its main aim is to improve road monitoring maximising vehicles flow and reducing waiting time at traffic lights. Each intersection is characterized by an Intersection Control Agent (ICA) that receive information from wireless sensors distributed along roads and determines the traffic flow model [14]. Another research [15] shows how it is possibile to equip vehicles with an electronic tag that can be mainly used for management traffic application. This solution would remove the sensor network infrastructure, since vehicles could easily exchange information each other. However, this approach is too invasive to be applied in a traffic management context. To improve real-time road monitoring, in addition to WSN, the fuzzy theory can be used. In fact, in the approach proposed in [16], a wireless sensor network collects traffic information [17] while fuzzy logic allows a dynamic and real-time traffic control. Based on number of detected vehicles by sensors, a profile is applied to manage the traffic flow. Performance are determined by calculating the Average Waiting Time Trip (AWTT) and results show wireless sensor network capabilities to quickly collect traffic information. In [18] is proposed a wireless sensor network to collect information on road intersections. These information can be transmitted to users that make requests about traffic. Indeed, the driver through his cell phone can receive traffic information of an interested area. The main goal of [19] is to design and implement an algorithm and, through several simulations, to evaluate road traffic management and control. The proposed system is able to detect ferrous objects in movement, e.g. vehicles, through a wireless sensors network. Furthermore, it is able to calculate the appropriate duration of green and red time of traffic lights in a road intersection. In this way, the system should help to solve traffic congestion problems. In any case, it is necessary to use many sensor nodes to increase system accuracy.

3.3. Video transmission problems over IEEE 802.15.4/ZigBee

In order to adapt video signal to the available bandwidth provided by IEEE 802.15.4/ZigBee protocol [7], it is really important to choose the best video coding approach. Surely IEEE 802.15.4/ZigBee is not a high performance protocol, but it is possible to transfer audio and video through an appropriate compression algorithm. Also, it is necessary to consider other characteristics in terms of video resolution and number of frames per second [20]. These specifications are used to adapt video signal to the available bandwidth (250 kbps) for each of sixteen channels in the ISM band (2.4 GHz).

3.3.1. Video compression

Video compression technologies [21] are intended to reduce and remove redundant information. In this way, digital video file can be transmitted over a network and stored on computer hard drives more efficiently. With effective compression techniques, it is possible to obtain a considerable file size reduction with minimal effects on image quality. However, it is possible that images quality is compromised if file size is further reduced increasing the compression level. There are various compression technologies, both proprietary and industrial standards. Standards are important to ensure compatibility and interoperability. They are particularly useful for video compressions because a video can be used for different purposes. For example, in many surveillance applications, it must be viewable even many years after the registration date. Among various surveillance systems [22], end users can choose among different suppliers, rather than being tied to one vendor. Motion JPEG [23], MPEG-4 [24] and H.264 [25] are most used algorithms in video surveillance. H.264 is the most recent and effective video compression standard. The compression process applies an algorithm to the source video in order to create a compressed file ready for transmission or storage. When the compressed file is played, an inverse algorithm is applied to produce a video with the original content. Time required to compress, transmit, decompress and display a file is called latency. An algorithm pair used together determine a video codec (encoder/decoder). Video codecs of different standards are generally not compatible each other. For example, an MPEG-4 decoder can not be used with a H.264 encoder, simply because one of two algorithms is not able to properly decode the output of the other. However, it is possible to implement multiple algorithms in the same software or hardware in order to allow different formats coexistence.

3.3.2. Image compression Vs. video compression

Compression standards use different methods and have various transmission rate, quality and latency. Compression algorithms are divided into two types: images compression and videos compression. Image compression uses "intra - frame" encryption technology [26]. Data is reduced within an image frame removing unnecessary information that may not be visible by human eye. In fact, human eye is more sensitive to luminance characteristics respect to chrominance characteristics. Motion JPEG is a typical example of this type compression standard. In a Motion JPEG sequence images are compressed or encoded as a single JPEG image.

Figure 4. Motion JPEG

Video compression algorithms, like MPEG-4 and H.264, use inter-frame prediction [27] to reduce video data from a frames series. Furthermore, there are techniques, such as differential encoding [28], where each frame is compared with a reference frame; pixels are encoded only if they have been modified respect to the reference frame.

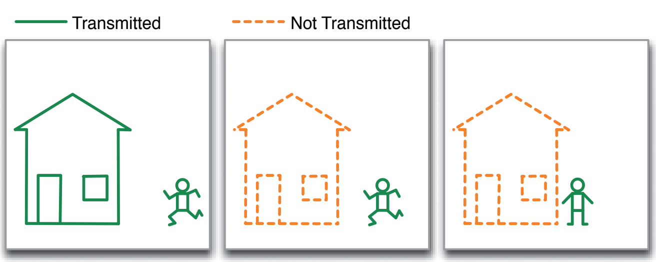


Figure 5. Differential Encoding

3.3.3. Motion JPEG

Motion JPEG or M-JPEG is a digital video sequence that consists of individual JPEG images. When 16 or more image frames per second are displayed, images are in motion perceived. Full motion videos are collected at 30 frames per second (NTSC) or 25 frames per second (PAL). One of Motion JPEG advantages is that various images of a video sequence can have the same quality, which varies depending on chosen compression level for the network camera or video encoder. If compression level is greater, then, image quality and file size will be smaller. Since there are no links between frames, a Motion JPEG video is "solid", i.e. if during transmission a frame is lost, the rest of the video will not be compromised. This format is characterized by a wide compatibility and it is prevalent in many applications. In particular, it is used when individual frames of a video sequence are needed and when there are reduced transmission rate, usually 5 or less frames per second. Main disadvantage of Motion JPEG is that it does not use video compression techniques to reduce data, since it consists of a complete images series. Result is a relatively high bit transmission rate or a low compression ratio to obtain performance comparable with to MPEG-4 and H.264.

3.3.4. MPEG-4

Usually, in video surveillance applications MPEG-4 Part 2 standard is used, also known as MPEG-4 Visual. Like other standards, MPEG license is purchasable, then users must pay a license fee associated with each monitoring station. MPEG-4 is used in applications characterized by limited bandwidth or requiring high images quality.

3.3.5. H.264

H.264 is used in high data rate and high resolution applications, (highways and airports) where 30/25 frames (NTSC / PAL) per second are needed. In fact, in these contexts, where it is necessary bandwidth and required storage space reduction, H.264 can offer the most significant advantages. This standard is also intended to accelerate network cameras diffusion with megapixels resolution because the efficient compression technology is able to reduce large files size and bit transmission rate without compromising images quality. However, H.264 requires the deployment of high performance network cameras and monitoring stations. Figure 6 shows performance of video encoding techniques described. Considering a sample video sequence, H.264 standard generates up to 50% bits per second less than an encoder which supports MPEG-4 with motion compensation. Furthermore, it is 3 times more efficient than MPEG-4 without motion compensation and at least 6 times more efficient than Motion JPEG.

4. System model

4.1. Introduction

As said, some problems remained open challenges such as the system ability to interact in harsh environments ensuring QoS for video monitoring traffic flows, and network predictability and reliability. The main goal of this work is to resolve these management issues through a fuzzy logic controller, highlighting improvements in terms of reduced packet loss and an increased number of packets successfully received by destination nodes. A network controller manages sampling period of each sensor node at run time, in order to maintain deadline miss ratio, associated with real time traffic flows, inside a desired level. In other words, the network controller dynamically determines the sampling period based on current value of deadline miss using apposite membership functions as provided by Fuzzy logic [10].

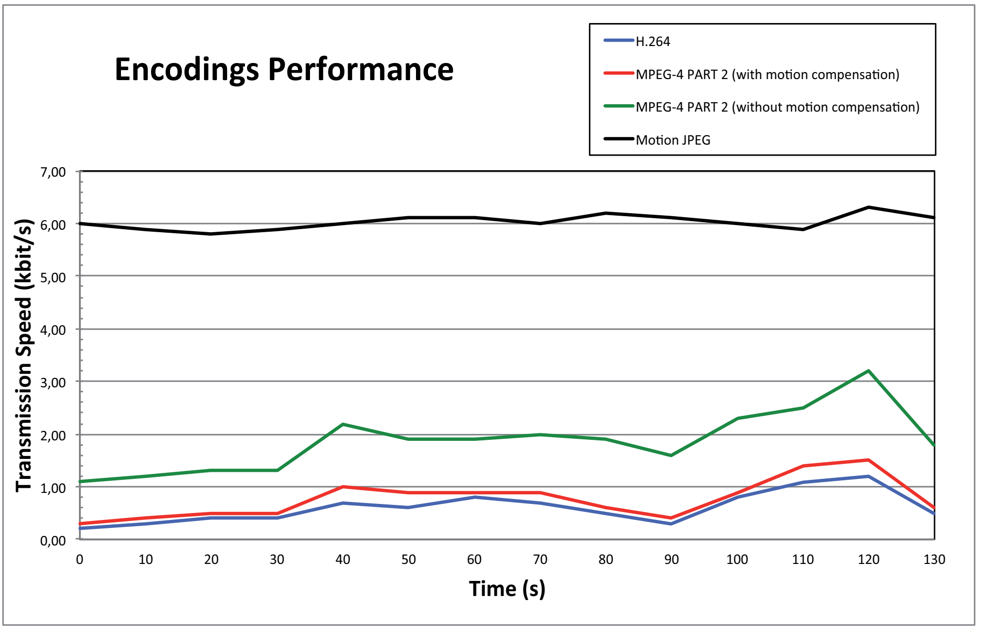


Figure 6. Performance of video encoding techniques

4.2. Network architecture

A road traffic monitoring system, working in adverse environmental conditions, needs the coexistence, in a WSN (based on IEEE 802.15.4), of several devices, which working together, provide a correct road traffic analysis. Figure 7 shows the system architecture here proposed. It integrates sensors connected to cameras with several magnetic sensor to reveal traffic flows.

The circled area in figure represents a cluster in which, when high traffic volume is detected, more monitoring cameras are dynamically activated. Surveillance cameras have the following characteristics to satisfy the available bandwidth (250 kbps) for each channel:

- Resolution 640 x 480
- 15 frames/s
- Medium Quality
- H.264 Encoding

Magnetic sensors are RFD nodes (included in black nodes in Figure 7) that, based on magnetic field distortion caused by the presence of ferrous objects like cars, provide basic information for traffic volume estimation. Subsequently, they send data detected to their cluster coordinator node (FFD). Its main task is to transmit data to and from other devices especially to the First Pan Coordinator (FPC), which is able to store and process network information and send, in case of high traffic volume detected, an activation message to "sleeping" nodes using the IEEE 802.15.4/ZigBee protocol. WSN functioning is appropriately and dynamically adapted to critical conditions in terms of congestion increase in order to optimize information quality. When a critical situation needs a more accurate monitoring, the

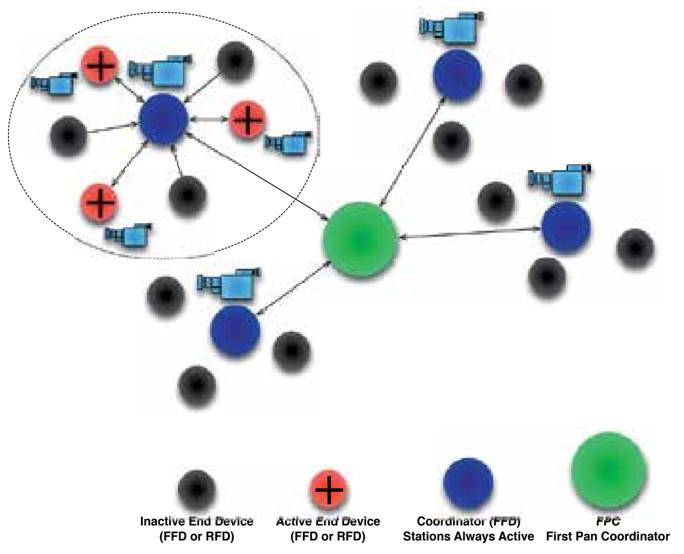


Figure 7. System Architecture

system here proposed activates more video control nodes based on traffic volume information detected. The increase of nodes in the network can cause a workload increase and then higher delay and packet loss. It is necessary to use a QoS management paradigm to ensure WSN flexibility, adaptability and scalability, considering network conditions changes. To this end, two basic features have been combined:

- smart network manager to optimize power consumption using a "low rate" protocol (IEEE 802.15.4/ZigBee)
- dynamic management of critical situations (also unexpected) which may lead network overloading.

4.2.1. Magnetic sensors

Magnetic field detection has enjoyed a significant growth thanks to several application fields: magnetic sensors can be used to detect the presence, strength or direction of magnetic field distortion caused not only by earth, but also by vehicles [30]. These sensors can measure these properties without physical contact and this is the reason why they have become so important for industrial control systems. We need to consider that magnetic sensors are never used to measure the magnetic field. Usually they are used to evaluate parameters like speed or vehicles presence. These parameters can not be directly calculated, but can be extracted based on magnetic field variation and distortion. Conventional sensors (temperature or pressure), can directly convert the desired parameter to a proportional voltage or current output. Using a magnetic sensor, it is necessary to measure the magnetic field variation and then process the measured signal in order to obtain the desired output. Figure 8 shows the difference between conventional sensors and magnetic sensors. Figure 9 show how a ferrous object, like a car, can create a local distortion. Vehicle detection applications can take different forms. A single axis sensor can detect vehicle presence. Magnetic distortion caused by a large ferrous object metals, like a car, can be modeled as a set of many magnetic dipoles. Figure 10 clarifies how a

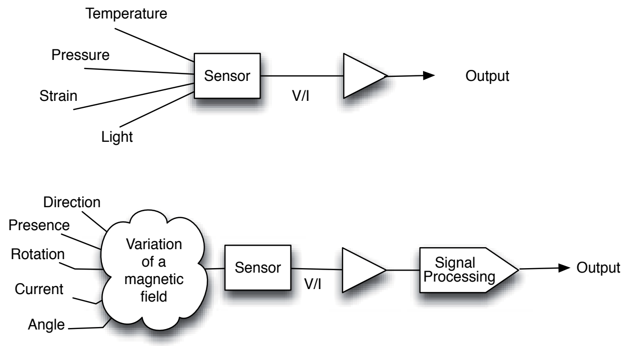


Figure 8. Difference between conventional sensors and magnetic sensors

uniform magnetic field is "distorted" by the presence of a ferrous object.

4.2.2. Vehicle classification

Magnetic distortions can be used to classify different types of vehicles [31, 32]. When a vehicle passes over the magnetic sensor, it will detect all different dipole moments. The field variation is a real "magnetic" signature of the vehicle. A three-axis magnetic sensor, positioned on the traffic lane, will provide a rich output signal related to vehicles passing over its area. Figure 11 shows a three-axis magnetic sensor output after detection of two vehicles (a pick up and a

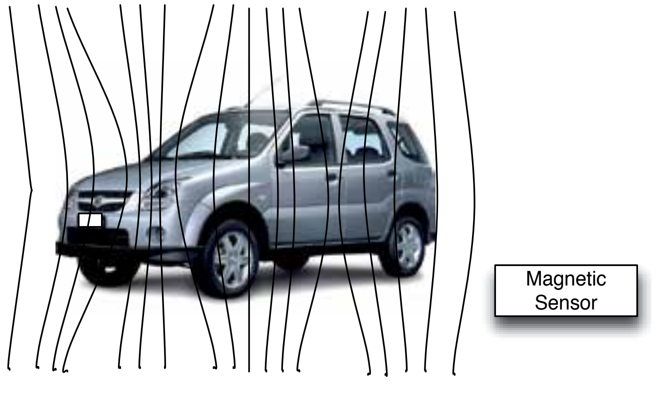


Figure 9. Magnetic field distortion caused by a ferrous object

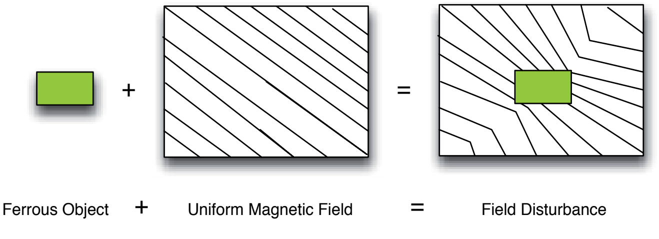


Figure 10. Magnetic field distorted

sedan respectively). The four curves represent the magnetic field variation of X, Y and Z axis respectively and amplitude caused by the vehicle proceedings towards the south. Vehicle type can be classified based on these variations using pattern recognition or matching algorithms. Magnetometer output curves in Figure 11, reveal how vehicles vary Earth's magnetic field. The greatest variation, in each curve, occurs when the engine block is exactly opposed to the sensor.

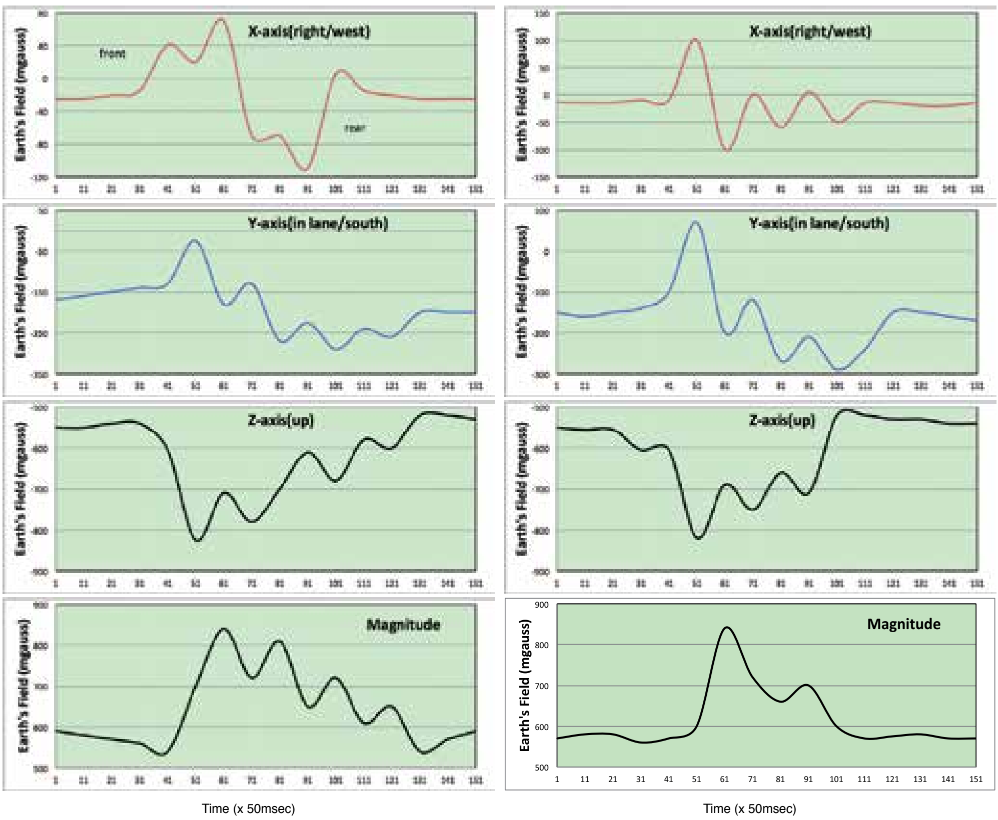


Figure 11. Vehicle classification

4.2.3. Vehicle direction and presence

Detection of vehicle presence and direction [33], does not require a high degree of detail in terms of magnetic field distortion. Example in Figure 12 evaluates the detection of a car driven at a distance from sensor of 1 and 3 foot respectively.

Curves X, Y and Z obtained are shown in Figure 13. On the left results obtained by a sensor positioned at a distance of 1 foot with car travelling from north towards south are shown. On the right, instead, results obtained by a sensor positioned at a distance of 3 foot with car travelling from east toward ovest are shown. Each curve has two deviations: the first one represents the car traveling in direction of travel, while second one represents the car in reverse.

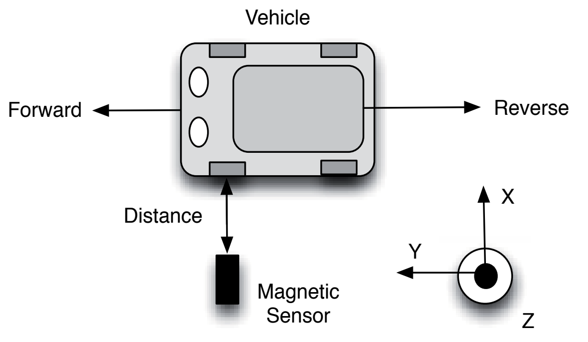


Figure 12. Vehicle Direction and presence

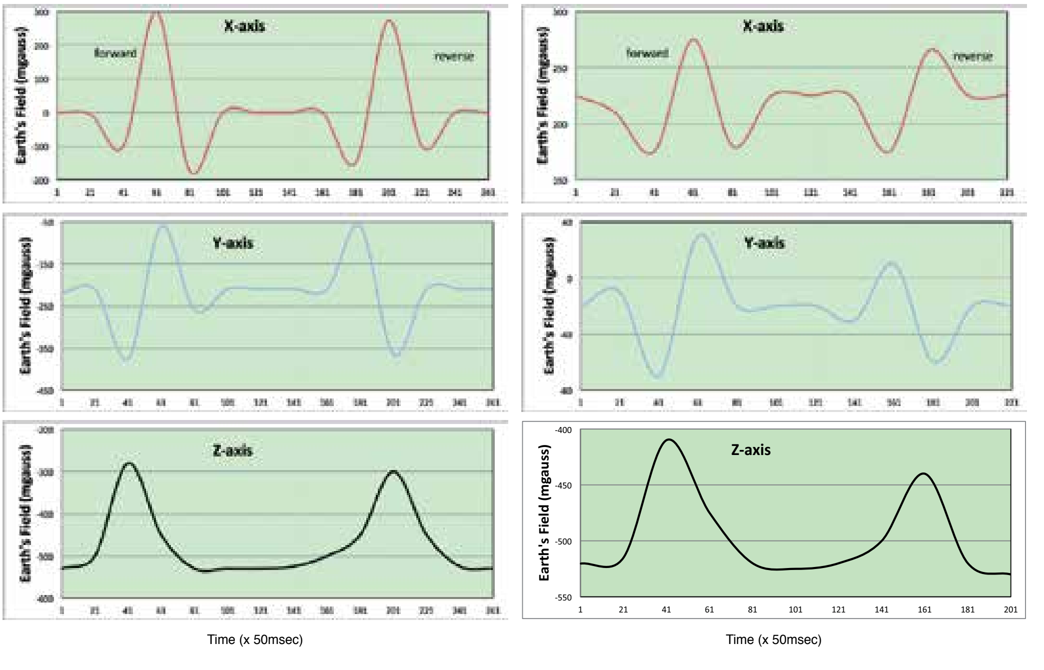


Figure 13. Direction Control

4.3. Dynamic video-monitoring algorithm

The approach here proposed (described through flow chart in figure 15), called Dynamic Video Monitoring Algorithm (DVMA), allows to activate more monitoring nodes in case of critical traffic situations detected by the WSN through magnetic sensors. Moreover, our approach controls the network load through a fuzzy controller whose task is to dynamically manage quality of service of data flows ensuring good performance. The DVMA evaluates, through magnetic sensors, the instantaneous magnetic field distortion value (VT). If VT measured value exceeds a threshold value, i.e. queue or traffic jam occur, and additional nodes are off, the FPC (green node in figure 7) sends an activation request to FFD nodes which, using the sleep/wakeup mechanism described in the IEEE 802.15.4 [8], forward the request to the sleeping RFD nodes. The idea is to forward an activation request

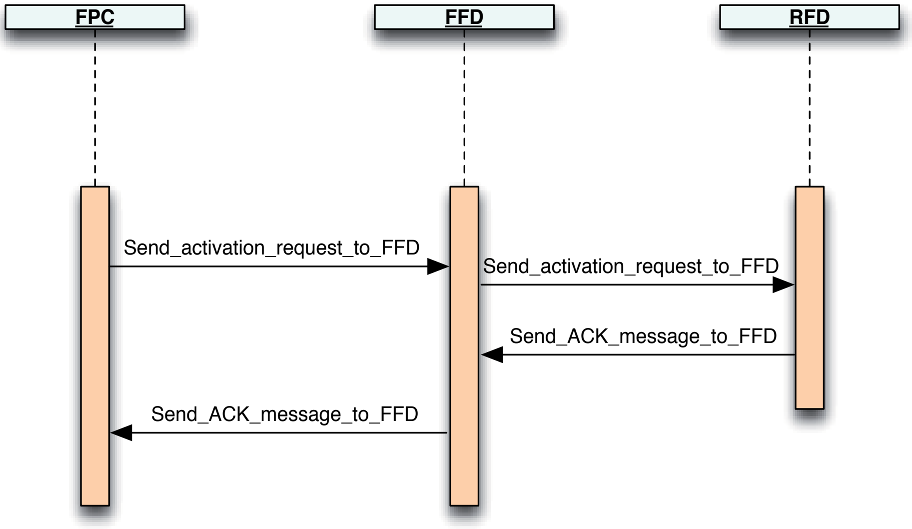


Figure 14. DVMA UML sequence diagram

setting to 1 the bit "High Traffic" introduced in the free reserved subfield of standard MAC control field frame as shown in figure 16. Sequence diagram in figure 14 shows DVMA functions. Threshold_inizialization() function defines the minimum traffic density. Beyond this limit, additional controls must be activated. According to magnetic field distortion, detected by RFD sensors, traffic volume trend can be analytically represented as shown in Figure 17. VT_acquisition() function acquires data collected by magnetic sensors placed along roads. Send_request_activation() function, sends an activation request from the First Pan Coordinator to "sleeping" RFD nodes through the FFD nodes (ZigBee routers) using 802.15.4/ZigBee standard protocol. Sleeping nodes wake up periodically, as described by standard [8], for channel control.

4.3.1. Fuzzy logic controller

The fuzzy logic controller manages network topology and workload. A similar approach has been applied in a context of WSAN [34] (Wireless Sensor Actuator Network). In our case, we have designed a fuzzy controller to dynamically change the sampling period of RFD nodes, determining the new sampling period (NST) based on two input values, as shown in figure 18:

- Deadline Miss Ratio Measured of packets (DMRM)
- Current Sampling Time (CST)

Based on predetermined "membership" functions [10], inputs are converted into "language" values: Positive Big (PB), Positive Small (PS), Zero (ZE), Negative Small (NS), Negative Big (NB). Subsequently, an inference mechanism, based on several IF-THEN rules, determines the output linguistic value representing the new sampling time NST (Positive Big, Positive Small, Zero, Negative Small, Negative Big). Figure 19 shows our inference mechanism functioning scheme. To better understand Figure 19, IF CST value is NS (Negative Small) and DMRM

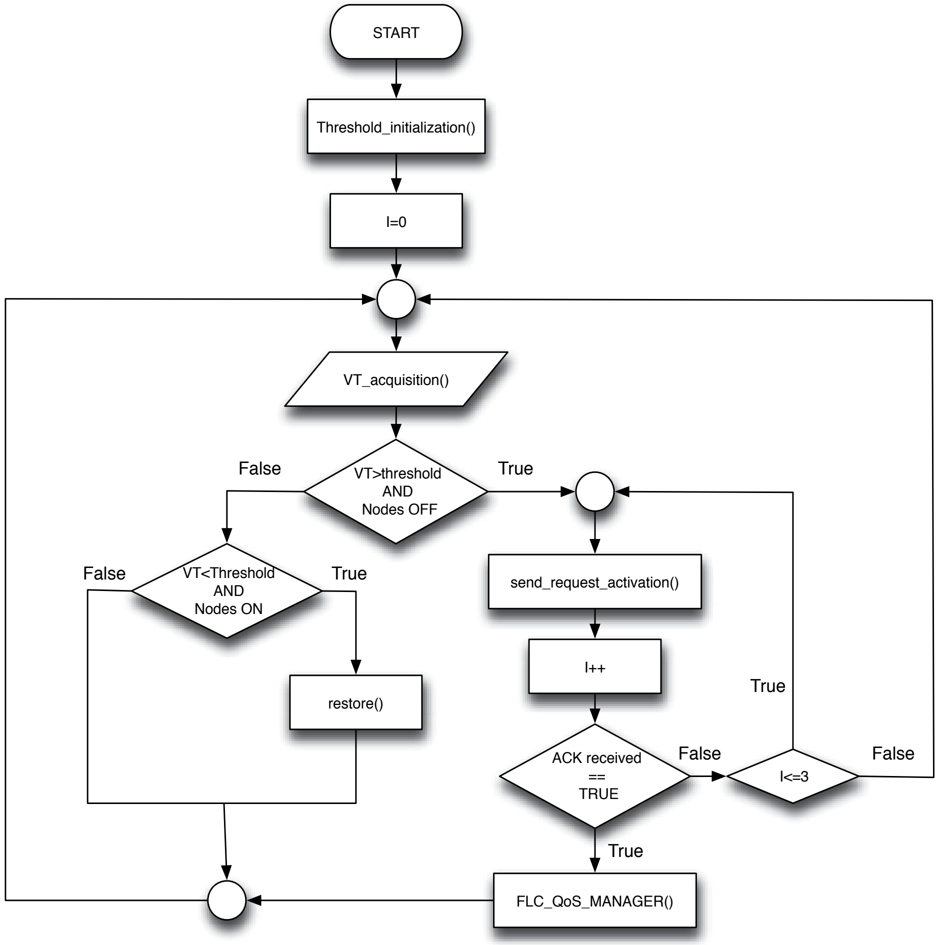


Figure 15. DVMA flow chart

Bits: 0-2	3	4	5	6	7	8	9	10-11	12-13	14-15
Frame Type	Security Enabled	Frame Pending	Add. Request	PAN ID Compress	Reserved	High Traffic	ADD Nodes request	Dest. Address. Mode	Frame Version	Source Address. Mode

Figure 16. Modified MAC control field

value is ZE (Zero), THEN NST value will be ZE (Zero). Finally, this value is defuzzified into a numeric value, which represents the new sampling period (NST) of RFD nodes. In our algorithm, for each variable, a range of value has been defined. Therefore the range has been divided in sub-ranges (called fuzzy sets). Established that Deadline Miss Ratio Measured (DMRM) can assume values between 0 and 1.25, this range can be divided into fuzzy sets as shown in Figure 20.

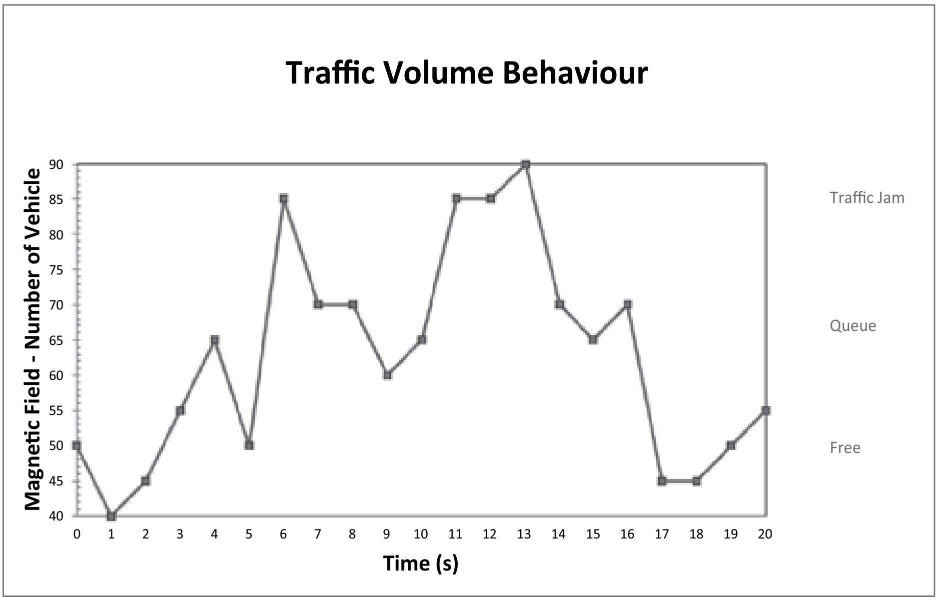


Figure 17. Traffic flows analytical behavior

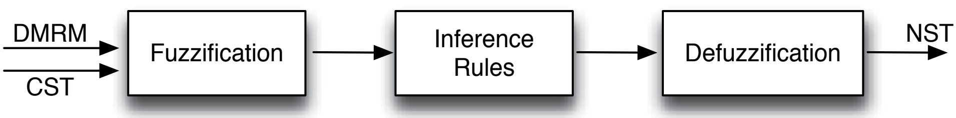


Figure 18. Fuzzy logic controller scheme

NST		**CST**				
		NB	**NS**	**ZE**	**PS**	**PB**
DMRM	**NB**	PB	PB	PB	PB	PS
	NS	PS	PS	PS	PS	ZE
	ZE	PS	ZE	ZE	ZE	NS
	PS	ZE	ZE	NS	NS	NB
	PB	NS	NS	NB	NB	NB

Figure 19. Inference mechanism example

In other words, if the value of DMRM is between 0 and 0.03, it will be fuzzyfied as Negative Big (NB). Similarly, Current Sampling Time (CST) can assume values inside a range divided in fuzzy sets too, as shown in Figure 21.

So, if Current Sampling Time has a value between 0 and 4, it will be fuzzyfied as Negative Big (NB). As seen previously, outputs are input functions. According to fuzzy logic, these functions are expressed through IF-THEN constructs. To better understand, the following construct (from figure 19) can be taken as model: **IF DMRM is NB And CST is ZE THEN**

NB	0	0.0015	0.03
NS	0.00265	0.0326	0.203
ZE	0.15	0.225	0.3
PS	0.25	0.5	0.7
PB	0.55	1.001	1.25

Figure 20. DMRM fuzzy sets

NB	0	2	4
NS	0	4	8
ZE	4	8	12
PS	8	12	16
PB	12	16	20

Figure 21. CST fuzzy sets

NST is PB. IF the DMRM value is Negative Big and CST value is Zero, THEN the NST value (New Sampling Time for FFD node) will be Positive Big.

5. Performance evaluation

To demonstrate benefits introduced by the DVMA, several simulations were carried out using:

- TrueTime [35]: it is a real-time simulation environment that allows co-simulation control tasks performed in real-time kernels.
- Simulink/Matlab has been used to test the DVMA in a WSN 802.15.4 based.
- OMNeT++: it ensure WSN evaluation [36]

TrueTime and Simulink/Matlab have been used to test DVMA performance in a single cluster network. OMNeT++ has been used to test performance of entire WSN system. The simulated network consists of a central FPC node. It covers 480 meters of road section in every direction thanks to FFD nodes located at a distance of about 50 meters from each other. Each FFD node (ZigBee router) provides a magnetic sensor, for traffic measures, and two cameras in "sleeping mode". These cameras will be activated only in case of real need. In other words, when magnetic sensors detect traffic volume increase.

The simulation campaign refers to periodic and aperiodic traffic flows. A periodic real-time traffic flow has regular arrival times (arrival time is equal to sampling time value). An aperiodic real-time traffic flow has irregular and unpredictable arrival times (it can be even one-shot). The first concerns video monitoring traffic flows, while aperiodic packets concerns data sent by magnetic sensors for traffic volume detection. During simulations, have been evaluated: number of packets received and lost for image traffic and network management traffic respectively. Packet size considered is 18 Kb, data-rate equal to 180 Kbps, Simulation Time equal to 30s. Figure 23 shows results obtained with fuzzy logic controller

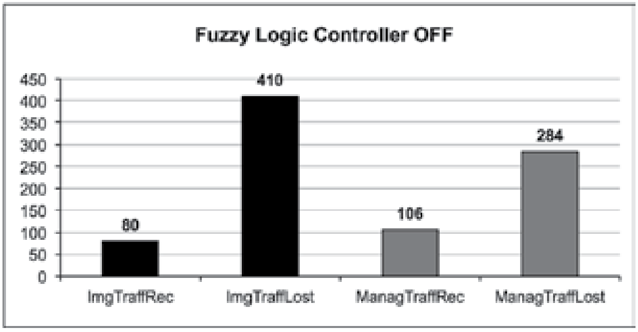


Figure 22. Performance obtained with controller OFF

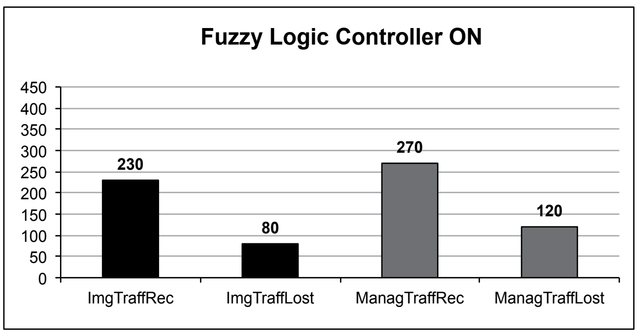


Figure 23. Performance obtained with controller ON

OFF while Figure 24 shows what obtained switching ON the controller. Results indicate a great improvement when the controller is ON: a higher number of image packets rightly received (230-270 Vs. 80-90) and, at the same time, a strong decrease of lost packets (80-120 Vs. 410-400). It is easy to observe that when the controller is OFF (Figure 22) the workload

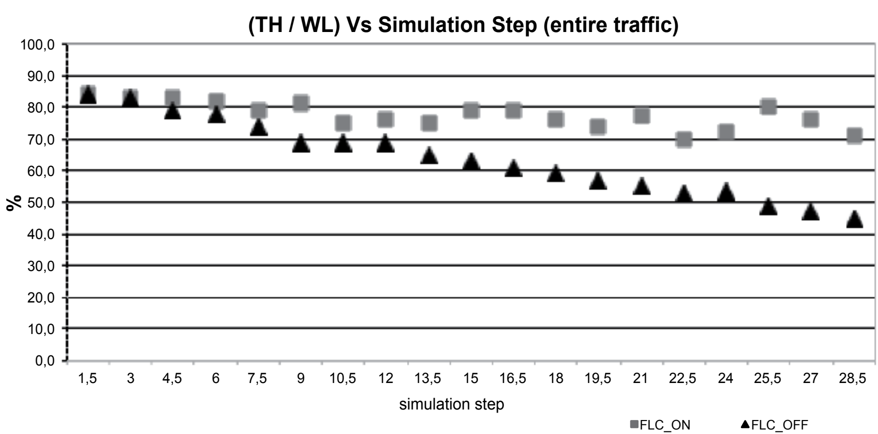


Figure 24. Th/Wl vs. simulation step (entire traffic)

related to periodic traffic is higher than the case with the controller ON. The fuzzy logic controller adjusting the periodic traffic (determining sampling times) reduces network traffic (workload) improving the throughput. In fact Figure 23 shows a number of periodic packet loss undoubtedly lower. Considering the aperiodic traffic, even if the network maintains the same workload, the DMVA improves performance because decreases general network workload.

We also analyzed network performance in terms of Throughput (Th) / Workload (Wl) (1). In other words:

$$Th/Wl = \frac{Number\_of\_Packets\_Successfully\_Received}{Number\_of\_Packets\_Generated} \quad (1)$$

Figure 25, Figure 26 and Figure 27 show how the fuzzy controller reacts to network degradations.

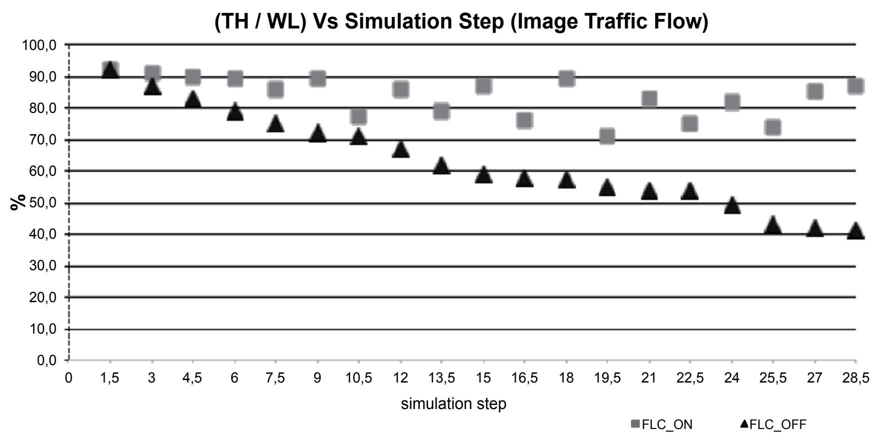


Figure 25. Th/Wl vs. simulation step (image traffic flow)

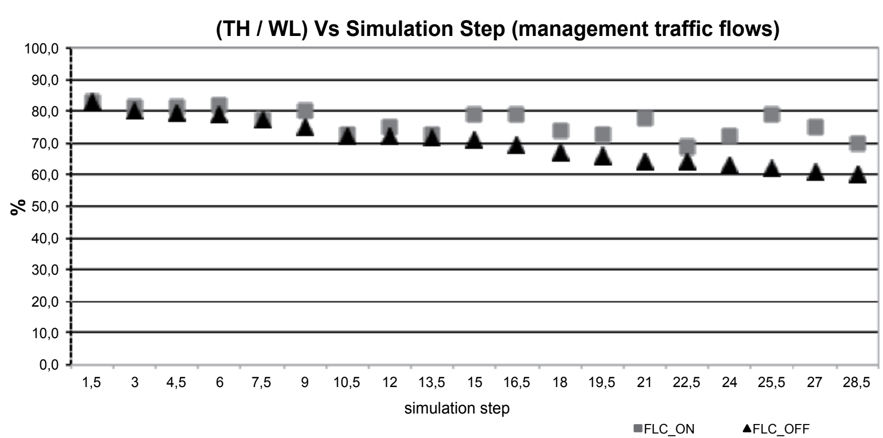


Figure 26. Th/Wl vs. simulation step (management traffic flow)

Figures 25,26,27 show how performance get worse when the controller is OFF. At each time window, in fact, the number of packets that correctly reach their destination, decreases till a minimum value measured of about 44% (total traffic sent), 41% (image traffic sent) and 59% (aperiodic traffic sent). Instead using the fuzzy logic controller, Th/Wl measured is equal to about 80% (total traffic sent), 76% (image traffic sent) and 82% (aperiodic traffic sent). The DVMA dynamically responds to an overloaded scenario. It has been simulated a generic situation adding and removing stations (due to fast road traffic variation) which generate additional traffic to the network. The network become able to prevent network collapse, while adding more monitoring nodes or when network conditions are close to saturation.

6. Conclusions and future works

This work discussed an algorithm to manage road traffic volume through a hybrid architecture that integrates magnetic sensors for vehicular traffic detection, with several cameras to ensure better monitoring. The possibility to activate more devices only in case of real need, ensures energy savings. Furthermore, using a fuzzy logic controller, we can manage easily and in a better way network workload changes. The controller works dynamically, based on QoS measured parameters and results are really promising. We are currently working on the implementation of the proposed architecture using COTS devices, with panels which indicate, along the monitored road section, the presence of heavy traffic and suggesting, where possible, alternative routes or estimated time to reach the desired destination. Moreover, our research will focus on real-time scheduling algorithms having as input data acquired through WSNs for road monitoring near traffic lights. In particular we are developing a mechanism to allow dynamic management of queues at traffic lights in order to give more green time to road sections with longest queues. In addition, we are studying a smart paradigm based on neural networks in order to predict, with mean square error as small as possible, vehicular traffic behavior based on values previously measured. As everybody knows, during the day streets are jammed in the morning (people go to work or accompany their children to school), lunchtime and late afternoon (work out). The Network Controller, after an accurate training phase, may cleverly monitor the established area in a soft way, dynamically determining the awakening of video surveillance cameras only when really necessary based on values detected by magnetic sensors and predictions determined by the neural algorithm. Another important research aspect concerns the study of power consumption within WSNs. Dynamic sampling times management allows us to assume a lower average power consumption of batteries used by the sensors. This leads an increase of batteries duration and network lifetime. Our goal is thus to measure energy savings that can be obtained through our approach and determine if the approach can be further improved by using a neural network system or a neuro-fuzzy combination.

Author details

M. Collotta, G. Pau, V.M. Salerno and G. Scatá
Faculty of Engineering, Architecture and Motor Science, Telematic Engineering Laboratory, Kore University, Enna, Italy

7. References

[1] Y. Wen, J.L. Pan and J.F. Le, "Survey on application of wireless sensor networks for traffic monitoring", Intenrational Conference on Transportation Engineering, pp.2079-2084, 2007

[2] C. Wenjie, C. Lifeng, C. Zhanglong, T. Shiliang, "A realtime Dynamic Traffic Control System Based on Wireless Sensor Network", 2005 INternational Conference Workshops on Parallel Processing, pp.258-264, 2005

[3] S.Y. Cheung, P. Varaiya, "Traffic Surveillance by Wireless Sensors Networks" IPh. D. Dissertation, University of california, ISSN 1055-1425, 2007

[4] S.Y. Cheung, S. Coleri, P. Varaiya, "Traffic Surveillance with Wireless Magnetic Sensors", ITS World Congress, 2005

[5] S.Coleri, S.Y. Cheung, P. Varaiya, "Sensor Networks for monitoring traffic" Allerton Conference on Communications, 2004

[6] S. Deshpand, "Adaptative low-bitrate streaming over IEEE 802.15.4 low rate wireless personal area networks (LR-WPAN) based on link quality indication", International Conference on Communication and Mobile Computing, pp. 863-868, 2006

[7] G. Pekhterverv et. al "Image Transmission over IEEE 802.15.4 and ZigBee Networks" IEEE International Symposium on Circuits and Systems, Vol.4, pp.3539-3543, 2005

[8] "IEEE 802.15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)" - IEEE standard for information technology, Part 15.4, 2006

[9] Bluetooth Specification Version 3.0 + HS; Bluetooth SIG 2009.

[10] L.A. Zadeh, "Toward extended fuzzy logic - A first step", Fuzzy Sets and Systems, Vol. 160, Issue 21, pp. 3175-3181, 2009

[11] M.Tubaishat, Qi Qi, Yi Shang, Hongchi Shi, "Wireless Sensor Based Traffic Light Control", 5th IEEE Consumer Communications and Networking Conference CCNC 2008, pp. 702-706, 2008.

[12] S.Y.Cheung, S.Coleri, B.Dundar, S.Ganesh, C.W.Tan, P.Varaiya, "Traffic measurement and vehicle classification with a single magnetic sensor", 84th Transportation Research Board Annual Meeting Technical report, pp. 173-181, 2005.

[13] M.Tubaishat, Y.Shang, H.Shi, "Adaptive Traffic Light Control with Wireless Sensor Networks", 4th IEEE Consumer Communications and Networking Conference (CCNC 2007), pp. 187-191, 2007.

[14] C.Wenjie, C.Lifeng, C.Zhanglong, T.Shillian, "A wireless sensor network for intelligent transportation system", First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS ’06), vol. 2, pp. 635-641, 2006.

[15] A.Guitton, A.Skrdasky and N.Trigoni, "Correlation based data dissemination in traffic monitoring sensor networks", Proceedings of the 2006 ACM CoNEXT conference (CoNEXT ’06), 2006.

[16] Z.Fuqiang, Y.Bo, C.Yitao, "Traffic Light Control for a single intersection based on Wireless Sensor Network", 9th International Conference on Electronic Measurement And Instruments (ICEMI 2009), pp.1040-1044, 2009.

[17] M.Wiering, J.Vreeken, J.Van Veenen, A.Koopman, "Simulation and optimization of traffic in a city", 2004 IEEE Intelligent Vehicles Symposium, pp. 453-458, 2004.

[18] Huang, Qigfeng, Ying, Zhang, "Dynamic balancing of push and pull in a distributed traffic information system", 4th IEEE Consumer Communications and Networking Conference (CCNC 2007), pp. 177-181, 2007.
[19] D.T.Dissanayke, S.M.R.Senanayake, H.K.D.W.M.M.R.Divarathne, B.G.I.T.Samaranayake, "Real-Time Dynamic Traffic Light Timing Adaptation Algorithm and Simulation Software", 2009 International Conference on Industrial and Information Systems (ICIIS 2009), pp.563-567, 2009.
[20] S. Misra, "A survey of multimedia streaming in wireless sensor networks", IEEE Communications Surveys And Tutorials, Vol. 10 Issue 4, pp. 18-39
[21] D.Marpe, H.Schwarz, T.Wiegand, S.Bosse, B.Bross, P.Helle, T.Hinz, H.Kirchhoffer, H.Lakshman, N.Tung; S.Oudin, M.Siekmann, K.Suhring, M.Winken, "Improved video compression technology and the emerging high efficiency video coding standard", 2011 IEEE International Conference on Consumer Electronics (ICCE 2011), pp. 52-56, 2011.
[22] M.H.Sedky, M.Moniri, C.C.Chibelushi, "Classification of smart video surveillance systems for commercial applications", IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS 2005), pp. 638-643, 2005.
[23] T.V.Dung, Q.N.Truong, "Quality Enhancement for Motion JPEG Using Temporal Redundancies", IEEE Transactions on Circuits and Systems for Video Technology, Volume: 18 , Issue: 5, pp. 609-619, 2008.
[24] K.Mamou, T.Zaharia, F.Preteux, "FAMC: The MPEG-4 standard for Animated Mesh Compression", 15th IEEE International Conference on Image Processing (ICIP 2008), pp. 2676-2679, 2008.
[25] Z.Xiaoli, T.Pin, Y.Shiqiang, K.Fei, "Computation Offloading for H.264 Video Encoder on Mobile Devices", Multiconference on Computational Engineering in Systems Applications (IMACS), pp. 1426-1430, 2006.
[26] X.Zhao, L.Zhang, S. Ma, W.Gao, "Rate-distortion optimized transform for intra-frame coding", IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2010), pp. 1414-1417, 2010.
[27] L.Xiaoming, Z.Debin, J.Xiangyang, W.Qiang, G.Wen, "A Fast Inter Frame Prediction Algorithm for Multi-View Video Coding", IEEE International Conference on Image Processing (ICIP 2007), pp. 417-420, 2007.
[28] O.Hwa-Yong, S.Erturk, C.Tae-Gyu, "Low Complexity Video Encoding with One-Bit Transform based Network-Driven Motion Estimation", IEEE Transactions on Consumer Electronics, Volume: 53 , Issue: 2, pp. 601-608, 2007.
[29] S.Bo, "Submacroblock motion compensation for fast down-scale transcoding of compressed video", IEEE Transactions on Circuits and Systems for Video Technology, Volume: 15 , Issue: 10, pp. 1291-1302, 2005.
[30] J. Lenz, "Magnetic sensors and their applications", IEEE Sensors Journal, Vol. 6, Issue 3, pp. 631-649, 2006
[31] S. Kaewkamnerd, "Vehicle classification based on magnetic sensor signal", IEEE International Conference on Information and Automation, pp. 935-939, 2010
[32] S. Kaewkamnerd, "Automatic vehicle classification using magnetic sensors", IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems, pp- 420-424, 2009

[33] Bong-gu Lee, "Algorithm for finding the moving direction of a vehicle using magnetic sensor", IEEE Symposium on Computational Intelligence in Control and Automation, pp. 74-79, 2011
[34] X. Feng, W. Zhao, Y. Sun and Y.C. Tian, "Fuzzy Logic Control Based QoS Management in Wireless Sensor and Actuator Networks", Sensors 2007, Vol. 7, No. 12, pp. 3179-3191, 2007.
[35] http://www.control.lth.se/truetime/
[36] http://www.omnetpp.org/

Wireless Sensors Network Application: A Decentralized Approach for Traffic Control and Management

Faisal Ahmed Al-Nasser and Magdi S. Mahmoud

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48212

1. Introduction

Wireless sensor networks (WSNs) are a significant technology attracting considerable research interest and have seen rapid growth due to the remarkable progress in microelectronics and electromechanical systems. Recent advances in wireless communications and electronics have enabled the development of low-cost, low-power and multi-functional sensors that are small in size and communicate in short distances. Cheap, smart sensors, networked through wireless links and deployed in large numbers, provide unprecedented opportunities for monitoring and controlling homes, cities, and the environment. In addition, networked sensors have a broad spectrum of applications in the defense area, generating new capabilities for reconnaissance and surveillance as well as other tactical applications [1,2]. Also, WSNs are comprised of numerous wireless sensor nodes that can sense light, temperature, sound, motion, etc. and wirelessly transmit them to a remote base station that aggregates the data and processes it locally or at another location.

In some applications, real-time, deadline violations that can occur in processing or transmission of collected data may result in some catastrophic events. Due to the necessity of time lines of computing in some of WSN applications, many valuable researches on real-time communications, power management and task scheduling have been done. Depending on the application, there may be a need to rapidly respond to sensor input. For instance, in a fire handling application, actions should be initiated on the event area as soon as possible. Moreover, the collected and delivered sensor data must still be valid at the time of action. Thus the issue of real-time communication is very important in WSN. Although several real time protocols have been proposed for mobile ad-hoc networks; due to differences between ad-hoc networks and WSN, it is not suitable to directly transplant those protocols into the design of WSN. In the large-scale sensor networks data integrity and reliability in addition to real-time communication is also an issue. The reliability of data is crucial to take an effective decision.

One important application is the traffic control and management system (TCMS) that is based on a wireless sensor network [12 – 16]. It gathers the traffic information and controls the traffic flow according to the incoming traffic data. Many traffic light systems operate on a timing mechanism, preset cycle time that changes the lights after a given interval. An intelligent traffic light system senses the presence or absence of vehicles then it controls the traffic lights accordingly. The idea behind intelligent traffic systems is that drivers will not spend unnecessary time waiting for the traffic lights to change which may lead them to some traffic violations and accidents when some drivers start to lose their patience.

An intelligent traffic system detects traffic in many different ways. The older system uses weight as a trigger mechanism. Current traffic systems react to motion to trigger the light changes based on the infrared object detector that picks up the presence of a car or some proximity switches. Then, a switch causes the lights to change. In order to accomplish this, algorithms are used to govern the actions of the traffic system. We need to understand the function of traffic signals so that we can improve driving habits by controlling the speed and the red light crossing in order to reduce the number of associated traffic accidents. The more the drivers know about the operation of traffic signals, the less frustrated they are going to be while waiting for the lights to change. Usually, in the intelligent traffic signal systems, the main aim is to reduce the cars waiting time at each signal and also to maximize the total number of cars that can cross an intersection safely with the green signal time.

In this chapter, the main goal is to gather the information of incoming vehicles via WSN for the decentralized deployed smart traffic light signals controllers along a the highway as shown in the Figure.1, that can do the following while maintaining fairness among the other traffic lights:

1. Intelligent traffic management system based on the queue length of cars in each signal side.
2. Optimize the following (trade-off):
   a. Minimize Average waiting time.
   b. Maximum possible service time (Green Duration).
   c. Minimize overall delay to vehicles.
   d. Maximize network capacity.
   e. Minimize the average number of cars of the same queue not passing from the same green duration.
   f. Minimize accident potential for all users.
3. Synchronizing the traffic lights phases that are placed in same directions along the highway to minimize the car stoppage, stop & start behaviour that build up queues and reduce the waiting time especially if the traffic in other sides is lower. For example, the intersections (A,B,C) we will maintain for the directions that cross all of them the green signal in a synchronized manner such that the queue length will be minimized and total travel time is minimum. We will apply the decentralized control approaches to handle that and the data collection will be the WSN.
4. Handling of the red light crossing violation to minimize the number of accidents by giving and alarm to the car which is about to start moving from the green light side.

5. The system can also give indication to the police traffic control room in case the road is blocked due heavy traffic to take an immediate action.

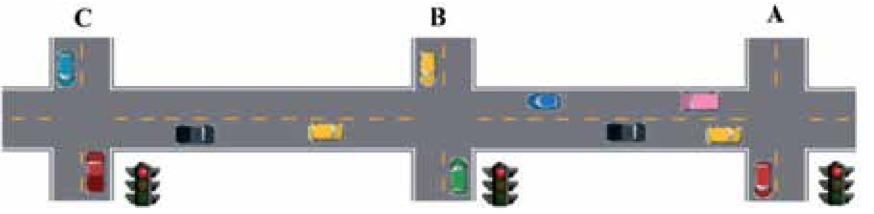


Figure 1. Traffic Signals Network

It is important to note that some of the objectives do conflict and a compromise may have to be made in the selection of objectives. However, some objectives can be met in tandem, for example minimizing delay to vehicles would also help to minimize total travel time and increase network throughput.

In planning and designing a traffic signal control system, one must first understand the applicable operational concepts related to signalized intersection control and signal-related special control. Signalized intersection control concepts include:

- Isolated intersection control - controls traffic without considering adjacent signalized intersections.
- Interchange and closely spaced intersection control - provides progressive traffic flow through two closely spaced intersections, such as interchanges. Control is typically done with a single traffic controller.
- Arterial intersection control (open network) - provides progressive traffic flow along the arterial. This is accomplished by coordination of the traffic signals.
- Closed network control - coordinates a group of adjacent signalized intersections.
- Areawide system control - treats all or a major portion of signals in a city (or metropolitan area) as a total system. Isolated, open- or closed-network concepts may control individual signals within this area.

Signal-related special control concepts include:

- High occupancy vehicle (HOV) priority systems.
- Preemption - Signal preemption for emergency vehicles, railroads, and drawbridges.
- Priority Systems - Traffic signal control strategies that assign priority for the movement of transit vehicles.
- Directional controls - Special controls designed to permit unbalanced lane flow on surface streets and changeable lane controls.
- Television monitoring.
- Over height vehicle control systems.

2. Problem statement and formulation

Controlling the traffic light intersection requires a prior knowledge of that intersection and the traffic load to be able set the proper parameters for the control algorithm, especially if the system used is not an intelligent system like time based traffic control. Basically most of the traffic signals [18] intersections have four directions queues, North (N), South (S), East (E) and West (W) (see Figure 2.). The other queues possibilities are North West (NW), South East (SE), East South (ES) and West North (WN) (see Figure 3). The model in Fig.3 simply shows that two directions can be open at the same time, for example, N and S direction will move then W and E at the same time because there is no turning in other directions like NW or SE. The other scenario is when we have the other directions NW; SE; ES and WN, then the control algorithm will be more complicated and more sensing elements are required. So, the main goal is to provide a controlling mechanism to minimize the waiting time for vehicles waiting in the red signal and maximize the service time for cars passing the green signal to avoid if possible the number of cars not able to pass from the first time or at least minimize this number. For simplicity, we will give a number for each queue q_i where $i = 1, \ldots\ldots 8$ for the following in order (*N, S, E, W, NW, SE, ES, WN*).

Figure 2. Basic Intersection

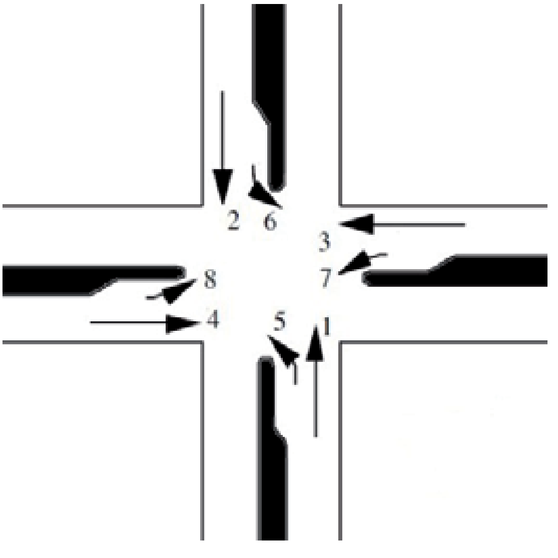


Figure 3. The Typical Intersection Directions

The intersection consists of four streets with 8 possible queues; assuming all right side movements are free and do not require a signal. The state equation for the continuous traffic flow process associated with any movement i that is sampled every t seconds, where time is indexed with the integer k, can be expressed by the current queue $q_i(k)$:

$$q_i(k+1) = q_i(k) + q_i^{in}(k) - q_i^{out}(k) + \phi_i^{in}(k) - \phi_i^{out}(k),\ i = 1,2,\ldots.8 \tag{1}$$

where $q_i^{in}(k)$ is the incoming new vehicles at time interval $[k-1,k]$ in link or queue i, $q_i^{out}(k)$ is the number of vehicles were able to pass the intersection during the green signal interval T_g from link or queue i, also T_g can be called as the control interval , $q_i(k+1)$ is the queue of vehicles were waiting for green signal to happen at time k, $\phi_i(k)$ represents the fluctuation between a parking lot and link i or the effects of any non-controlled intersection between any two intersections where $\phi_i^{in}(k)$ used for vehicles left the parking or came from non-controlled intersection and joined the traffic in the queue i and $\phi_i^{out}(k)$ for vehicles left the queue i and went for a parking or went into a sub road or what we call it non-controlled intersection. These disturbing flows can be considered either as disturbance or as known perturbations if they can be well measured or estimated. In case of these uncertainties or perturbations are unknown and can't be measured, then robust control system is needed.

The general discrete LTI state space representation is the following:

$$x(k+1) = A\,x(k) + B\,u(k) + F\,d(k) \tag{2}$$

$$y(k) = C\,x(k)$$

The state matrix **A** is practically considered as an identity matrix. The elements of the state vector x(k) represent the number of vehicles of each controlled link or in another word the queue length in that lane and the number of states is equal to the number of controlled links in the network. The second term of the state equation is the product of input matrix B and control input u where the vector u contains the green times of all stages. Matrix B can be constructed by the appropriate allocation of the combinations of saturation and turning rates. Their numerical values are the results of a corresponding controller at each cycle. The diagonal values of B are negative and represents the saturation flow and the product of $B_{ij}\,u_i$ where $i = j$; diagonal elements shows the outflow from link i. The other elements in B_{ij} where $i \neq j$ contains the turning rates from link i to link j. Naturally the number of states is equal to the number of controlled links in the network. The product B u(k) is arising from difference of in and out flow for the traffic in the link or queue i during the control interval. Each output inside of the network is a measured state (number of vehicles of the link i that makes the output equation simplified to, y(k) = x(k) and C = I. Finally, the traffic coming from non-controlled intersections or parking are considered as disturbance to the system in d(k).

Flow characteristics of traffic are fundamental in analyzing intersection delay or capacity. Vehicles occupy space and, for safety, require space between them. With vehicles moving continuously in a single lane, the number of vehicles passing a given point over time will depend on the average headway or the average arrival rate per unit time.

Two factors influence capacity at a signalized intersection:

- Conflicts occur when two vehicles attempt to occupy the same space at the same time. This requires allocation of right-of-way to one line of vehicles while the other line waits.
- The interruption of flow for the assignment of right-of-way introduces additional delay. Vehicles slow down to stop and are also delayed when again permitted to proceed.

These factors (interruption of flow, stopping, and starting delay) reduce capacity and increase delay at a signalized intersection as compared to free-flow operations. Vehicles that arrive during a red interval must stop and wait for a green indication and then start and proceed through the intersection. The delay as vehicles start moving is followed by a period of relatively constant flow. The Figure.4 illustrates the relation between the traffic flow and density (Fundamental diagram of traffic flow) and what can happen if the flow reaches the maximum and exceeds the critical density point which at the end leads to the jam density point where no vehicle will move.

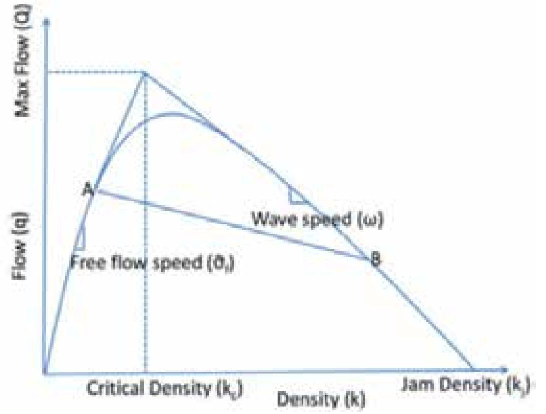


Figure 4. Traffic Flow Density Relation

Similar to the flow-density relationship, speed will be maximum, referred to as the free flow speed, and when the density is maximum, the speed will be zero. The simplest assumption is that this variation of speed with density is linear as shown by the solid line in Figure 5. Corresponding to the zero density, vehicles will be flowing with their desire speed, or free flow speed. When the density is jam density, the speed of the vehicles becomes zero. It is also possible to have non-linear relationships as shown by the dotted lines [9].

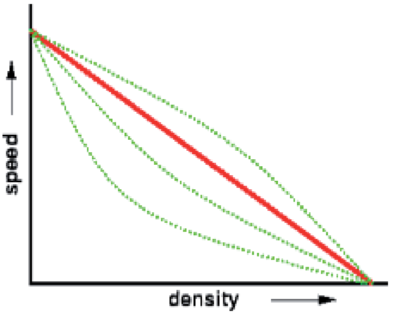


Figure 5. Speed-density diagram

The 3rd relationship is between the speed and flow where the flow is zero either because there are no vehicles or there are too many vehicles so that they cannot move. At maximum flow, the speed will be in between zero and free flow speed. This relationship is shown in Figure 6. The maximum flow Q_{max} occurs at speed certain speed which is the changing point in the parabola. Also, it is possible to have two different speeds for a given flow [10].

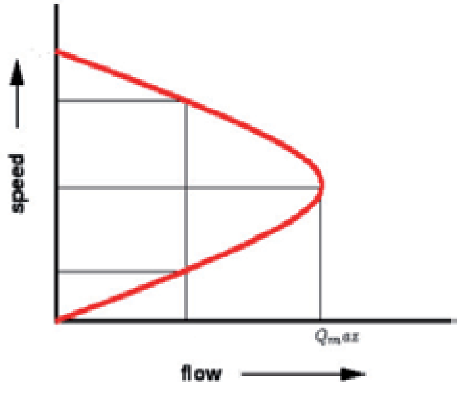


Figure 6. Speed and Flow

The diagrams shown in the relationship between speed-flow, speed-density, and flow-density are called the fundamental diagrams of traffic flow. They can be combined as shown in Figure 7.

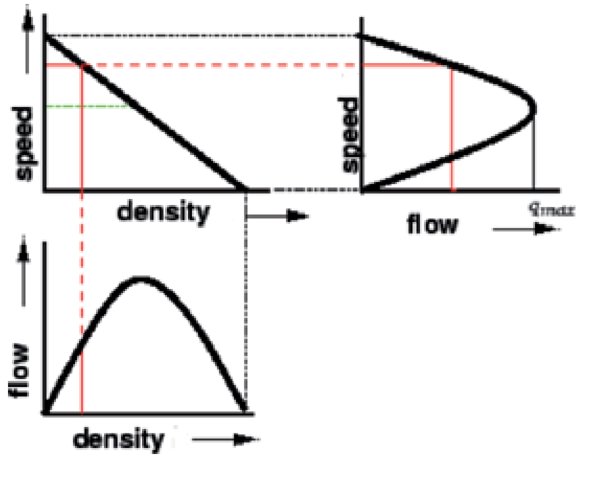


Figure 7. All Traffic Fundamental Relations

In case of detectors are not used, the simplest approach to model vehicle arrivals is to assume a uniform arrival. This will results in a deterministic, uniform arrival pattern which means constant time headway between all vehicles. However, this assumption is usually unrealistic, as vehicle arrivals typically follow a random process. Thus, a model that represents a random arrival process is needed and the most suitable one is the Poisson distribution with arrival rate of λ . In general, the car arrival is part of the queuing model (e.g. M/M/1 or M/G/1) which simulates the traffic signal operations. Basically the queue model is any service station with the following:

- One or multiple servers
- A waiting area or buffer

The time τ_n is interarrival time between cars *n* and *n+1* and it is a random variable. The traffic light system is following the stochastic process behaviour.

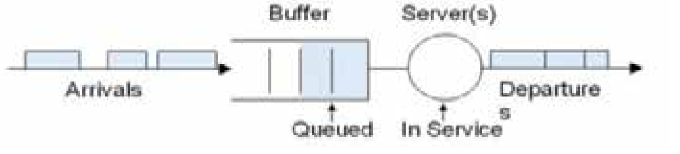


Figure 8. Basic Queue System

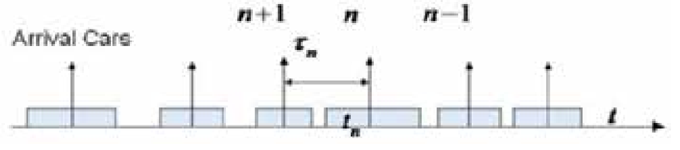


Figure 9. Arrivals Cars in time

3. Traffic signal phasing

The green signal period given for each side or combination of directions will be called phase (Figure.7) and the time will be T_g. The combination of phases can be called as Cycle where each phase or cycle must not exceed certain period to maintain the fairness for all direction in that intersection and it shall not be less than certain minimum. In all situations, the phases time shall not push the situation in that intersection to exceed the saturation level which will lead to traffic jam as we can see from Figure.4.

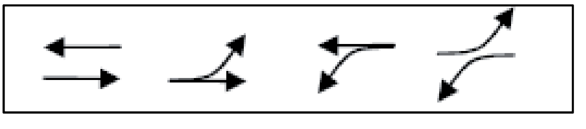


Figure 10. Different Phases for Traffic Signal Intersection (Examples)

Phasing reduces conflicts between traffic movements at signalized intersections. A phase may involve:

- One or more vehicular movements.
- One or more pedestrian crossing movements.
- A combination of vehicular and pedestrian movements.

The National Electrical Manufacturers Association (NEMA) has adopted and published precise nomenclature for defining the various signal phases to eliminate misunderstanding between manufacturers and purchasers [3]. Figure 5 illustrates a 4-phase sequence separating all vehicular conflicts. Holding the number of phases to a minimum generally improves operations. As the number of phases increases, cycle lengths and delays generally increase to provide sufficient green time to each phase. The goals of improving safety (by adding left-turn phases) and operations at a signalized intersection may conflict, particularly with pre-timed control. Operational efficiency at a signalized intersection, whether isolated or coordinated, depends largely on signal phasing versatility. Variable-sequence phasing or skip-phase capability proves particularly important to multiphase intersections where the number of change intervals and start-up delay associated with each phase can reduce efficiency considerably. Each set of stored timing plans has a distinct phase sequence.

Full-actuated traffic control illustrates variable-sequence phasing. In Figure 3, all approach lanes have detectors, using these detectors; actuated control skips phases with no traffic present and terminates certain movements when their traffic moves into the intersection. This capability produces a variation in the phasing sequence. The phasing options selected may be changed with the signal timing plan.

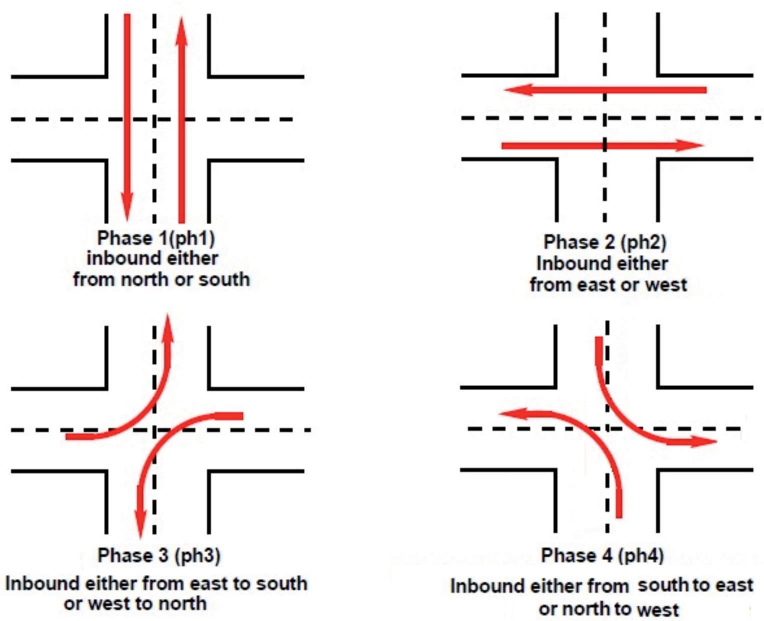


Figure 11. Example of Four Phases Intersection

4. Traffic detection at signalized intersection

Vehicle detectors (or in another word "Sensors") at a signalized intersection serve the singular purpose of informing the controller that a vehicle is (or is not) present on a particular approach to an intersection at any point in time. The controller uses this information to regulate the assignment of green time among competing movements at the intersection. The detector configuration is specified in terms of type, shape, length and setback from the stop line. These four parameters are all related to the physical installation of the detectors themselves. At signalized intersections with multilane approaches and actuated control, vehicle detectors are usually placed in each individual lane. Currently, the detectors across all lanes on a particular approach are linked together and are channeled to the same signal phase for controlling the phase duration. When any of the detectors detects a vehicle, the controller's gap-out timer resets and the phase (green) extends. Such a detection scheme makes it difficult to gap out a phase based on the desired headway or gap, especially when an approach has more than one lane. The unnecessary green extension directly affects the efficiency of signal operations, in which the extra green could be allocated to better serve other traffic movements. Ideally, a signal phase should terminate when a gap-out is reached for each lane individually [5]. The selection of the type of detector in signalized intersections is determined primarily by suitability for the intended purpose. The decision whether a particular detector is appropriate for a certain purpose depends on its operating characteristics, the cost, its adaptability to the particular application, and the location specific details of the installation requirements. Generally, vehicle detectors at signalized intersections are designed to sense either the presence of a waiting vehicle or the passage of a through vehicle.

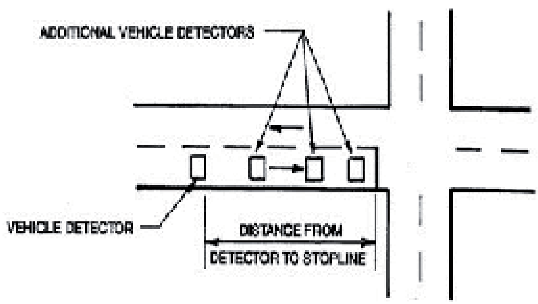


Figure 12. Multi Detectors Design

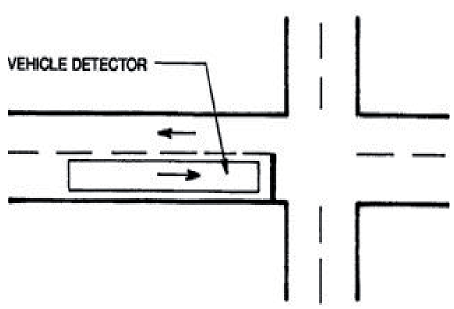


Figure 13. Long Presence Detector Design

Several types of detectors can be used and a wireless version for each type is possible for the data transfer. The following list shows these types in brief:

- **Inductive loop detectors (ILDs)** are the most commonly used type [5,6] of vehicle detection because of their reliability in reporting vehicle presence. However, loop detector installation can be expensive because of the physical connection required to connect the loops back to the traffic cabinet. Such connections are sometimes infeasible at locations such as bridges and ramps. Furthermore, saw-cut inductive loops are particularly sensitive to moisture and wire breaks associated with pavement failure.

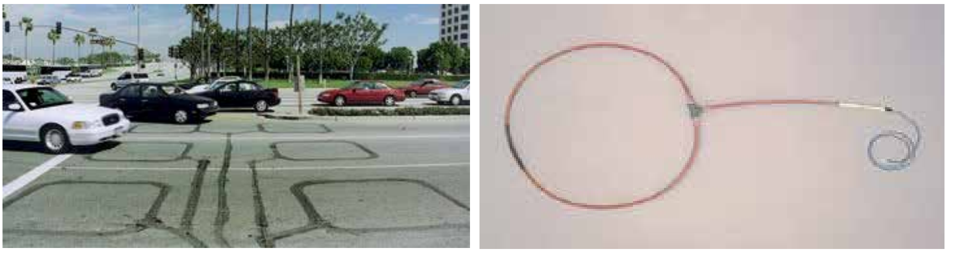


Figure 14. Inductive loop detectors (ILDs)

- **Compact wireless magnetometers** are a promising alternative to loops [5,6] because they require substantially less pavement cutting and no physical connection to a monitoring device.

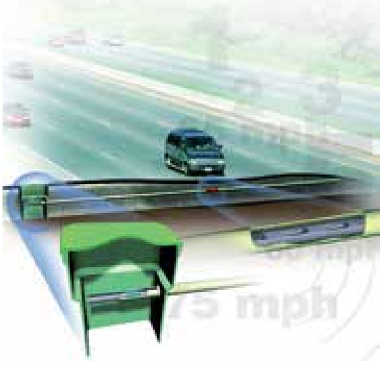


Figure 15. Compact wireless magnetometers

- **Video Imaging Vehicle Detection Systems (VIVDS)** have overcome some of the problems with loops such as traffic disruption and pavement degradation; but they have not been as accurate in all weather and light conditions as originally anticipated [7].

Figure 16. Video Imaging Vehicle Detection Systems (VIVDS)

- **Digital Wave Radar Detectors** technology to measure presence and speed of approaching vehicles in certain distance range based on the sensor features. It is well tested and very efficient.

Figure 17. Digital Wave Radar Detectors

- **IR detectors**: Infrared sensors are often used to detect stopped vehicles and also to detect pedestrians at pedestrian crossings.

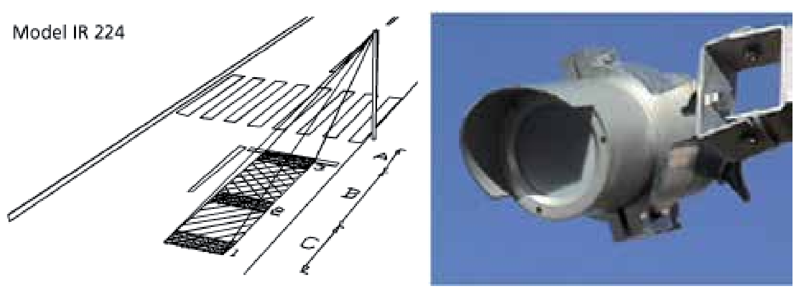


Figure 18. IR detectors

- **Ultrasonic Detectors:** transmit sound at 25 KHz to 50 KHz (depending on the manufacturer). These frequencies lie above the audible region. A portion of the transmitted energy is reflected from the road or vehicle surface into the receiver portion of the instrument and is processed to give vehicle passage and presence. A typical ultrasonic presence detector transmits ultrasonic energy in the form of pulses.
- **Passive Acoustic Detectors**: Vehicular traffic produces acoustic energy or audible sound from a variety of sources within the vehicle and from the interaction of the vehicle's tires with the road surface. Arrays of acoustic microphones are used to pick-up these sounds from a focused area within a lane on a roadway. When a vehicle passes through the detection zone, the signal-processing algorithm detects an increase in sound energy and a vehicle presence signal is generated. When the vehicle leaves the detection zone, the sound energy decreases below the detection threshold and the vehicle presence signal is terminated.
- **Combined technologies**: Above ground detectors using several different detector technologies are especially useful on highways or in tunnels to provide a wide variety of accurate detection parameters and classification information from a single location.

Also, the classification of detectors can be based on the function [8] of these detectors and the physical location that it will be placed on. Based on the functions of vehicle detection at signalized intersections, the main types of vehicle detections at signalized intersections are listed as follows:

1. **Advanced detection**: they are located in advance of the stop line. They are only used to detect moving vehicles. The common function of advance detection is to provide dilemma zone protection on high-speed approaches to signalized intersections. The dilemma zone is an area on the approach to a traffic signal whereby there are varied responses by drivers to the onset of the yellow signal. There is a different length dilemma zone for each approach speed. Also, these can be used for counting the arrival cars and send the information to intersection controller prior to arrival to signal for better selection of phase and green timing.
2. **Stop line detection**: are the most common and are so named because they are located at the stop line. Stop line detectors require greater sensitivity as slow moving or stopped vehicles must be detected. The location of the loop in relation to the stop line must be such as to ensure that a vehicle's normal stopping position is in the detection zone. In addition the detector must have sufficient memory time to monitor waiting traffic even under conditions of extreme congestion.
3. **Left & Right turn detection:** these will detect the cars going for the right side.
4. **Counting detection:** These traffic counting detectors can be used to count traffic in individual lines, lanes in a particular direction simultaneously, or all lanes in both directions continuously. As the number of lanes counted by a single detector increases, the accuracy of the count decreases as multiple vehicles can occupy the same detector at the same time.
5. **Violation detection**: Violation detectors are installed in conjunction with a red signal violation camera and flash unit to enable red signal traffic violations to be detected. If a vehicle passes over one of the detectors while facing a red signal the camera and flash are activated.
6. **Truck Detection**: The truck detector would be an added detector and would be placed much farther from the intersection. Its purpose would be to grant a green extension that would carry the truck to the normal detector location where it would also get an additional green extension. Therefore, if the "last vehicle' arriving at an intersection during the green interval is a truck, it will get dilemma zone protection as well as a green extension. If the last vehicle over the normal detector is a car, but there is a truck following that has actuated the truck detector, the truck will have sufficient green time to also reach the intersection. These detectors should have the ability to detect the specific vehicles such as trucks or buses with a high accuracy.

The improper placement of detectors can increase the lost time per phase and therefore the total lost time per signal cycle. This lost effective green time leads to an increased individual and total delay for vehicles using the intersection. For a signal installation which is operating near capacity (i.e. with $0.85 < v/c < 0.95$) it is highly likely that the green time on a particular phase will be terminated by a vehicle interval extension gapping out rather than by the green time maxing out. Studies have shown that at efficiently designed intersections, that is intersections operating near capacity where the mean service rate is greater than the mean arrival rate, the signal timings can be set such that 90% of the signal terminations will involve a gap out and only 10% of the signal terminations will involve a max out (a max out corresponds to a cycle failure). Maxing out of the signal phase, and hence a cycle failure, will be caused by random fluctuations in the arrival rate and can be predicted by employing a stochastic delay model.

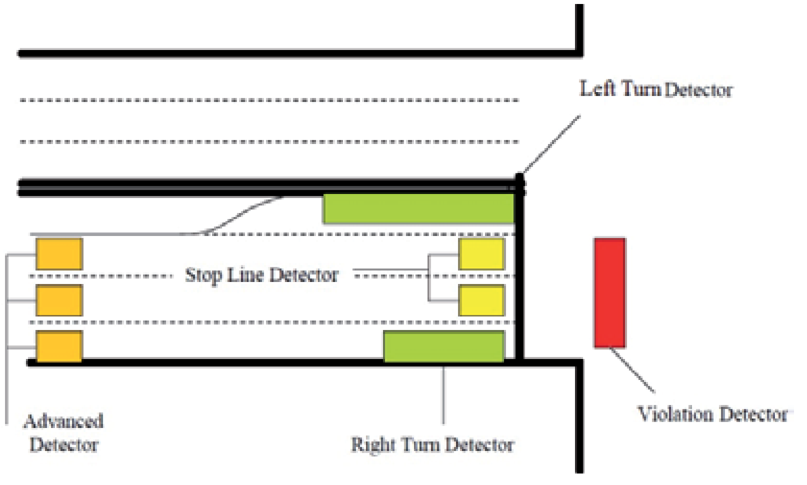


Figure 19. General Layout of Vehicle Detectors at Signalized Intersection

5. Signal control strategies

From the traffic engineering point of view, several different strategies are employed for the control of traffic signals ranging from non-actuated fixed timed to fully traffic responsive volume-density control. Fixed time, also known as pre-timed, controllers are of limited use for isolated intersections and require no traffic detection. Traffic actuated controllers are by far the most common control method used for isolated intersections. These traffic actuated controllers can range from semi-actuated, to fully traffic actuated, to traffic responsive volume-density. All traffic actuated controllers require vehicle detection on all legs of the intersection for efficient operation.

The general operation of traffic actuated controllers is described as follows:

- The green time for each phase is determined by the volume of traffic on the corresponding street and may vary from cycle to cycle. A maximum green time is predetermined and set within the controller.
- The request for green time is placed by a vehicle detector actuation during the red phase of a conflicting traffic movement. The minimum initial green time available is predetermined and set within the controller. This minimum initial green time is usually set to be adequate for the number of vehicles waiting between the stop line and the vehicle detector.

- Each additional vehicle which actuates the detector during the green phase calls for a vehicle green interval extension of a predetermined length. This extends the minimum green time up to the maximum green time set in the controller. Figure 20. Illustrates the vehicle interval extension process. If a gap in between vehicles occurs which is larger than this preset vehicle interval extension, and a call has been placed by an opposing phase, then the controller will 'gap out' and the green will be terminated for that phase. If enough traffic is present for the controller to reach the maximum green time then the controller will 'max out' and the green time will be terminated.

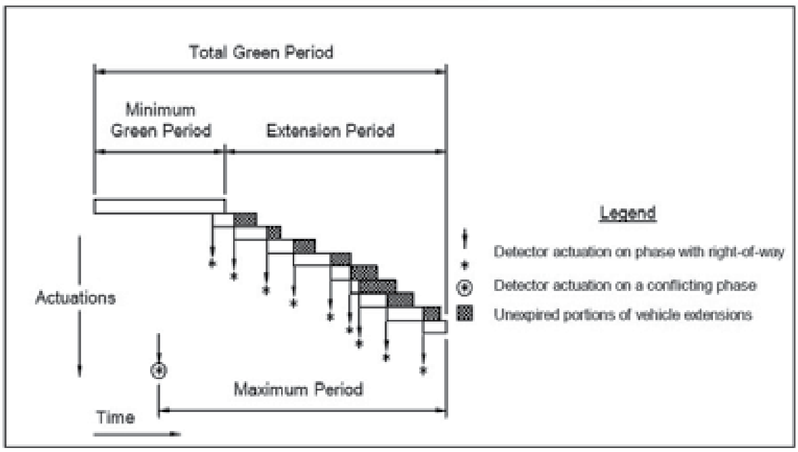


Figure 20. Actuated Controller Intervals

Some variations in the operation specific types of traffic actuated controllers is described as follows:

5.1. Semi-actuated controllers

- Semi-actuated controllers are used at intersections where the minor street has traffic volumes significantly lower than the major street. The priority of operation is to minimize the interruption of traffic on the major street while still providing adequate service to the minor street.
- Vehicle detectors are required only on the minor street. The detectors will input a call for green time as well as calls for vehicle interval extensions up to a preset maximum limit.
- The major street has a preset recall to its green phase. No detector call is required and the green will always revert to major phase when the minor street has been serviced. The major street will have a preset minimum green time, and will continue to rest in green on that phase until a call has been placed by the minor street.

5.2. Fully-actuated controllers

- Fully-actuated controllers are used where both streets at an intersection have relatively equal volumes. These controllers are particularly efficient where the traffic flows are

sporadic and uneven. The priority of operation is to minimize the total delay by minimizing stops on all phases.
- Vehicle detectors are required on all legs of the intersection. These detectors will input calls for initial minimum green time as well as vehicle interval extensions.
- Any or all phases of the signal can be set for automatic recall to green which will give the corresponding phase the minimum green time without a vehicle actuation. If all phases have automatic recall set then the signal will cycle through all the phases, giving each phase a minimum green time, even if no traffic is present on a particular leg. If recall is not set on any phase then the signal will service only those phases that have traffic actuations and skip the others. In the absence of any traffic, such as nighttime operation, the controller can either rest on the last served green phase or rest on red on all phases. If recall is set on only one phase the signal will revert to green on that phase once during every cycle. In the absence of traffic the signal will rest on green on this phase. This is a common setting for many intersections as one of the legs is usually considered slightly more major than the others.

5.3. Volume-density controllers

- Volume-density controllers are similar in operation to fully actuated controllers; however, contain more advance features for analyzing the traffic volumes on the green phase being served and the traffic density on the red phase being held. This information is then processed and the timing patterns altered for a more efficient operation. These controllers are the most efficient means of operation signals at isolated intersections.
- The most important feature of volume-density controllers is the ability to reduce the green vehicle extension interval depending on the density of opposing traffic. As the measured density of the opposing traffic increases the vehicle extension interval for green time is reduced linearly (known as gap reduction) to some preset minimum extension time.
- The controller has the ability to increase the minimum green time depending on the number of vehicles queued behind the stop line.
- In order to function properly these controllers must obtain information early enough to react to the fluctuating traffic patterns. Detectors must be place well in advance of the stop lines for such information to be useful.
- A special version of the volume-density controller, known as a 'Modified-Density Controller', has many of its features but requires less information from the intersection. Traffic flow statistics are obtained from the detector actuations of the previous cycle with the assumption that the present cycle being served has the same characteristics as the cycle preceding it. This may be acceptable for situation where traffic flow is relatively deterministic, however, is not efficient where intersection have highly unpredictable and random flows.

5.4. Fixed-time controllers

- Also we call it Pre-timed, in which we program fix phases sequence & timing based on data collected for peak hours for certain time. It may do a good job for less crowd intersections and also lower cost but it has the major disadvantage of being unable to adapt to changing conditions in real-time. At best they can be manually updated using accumulated traffic data however this is a time-consuming exercise and could not be carried out to allow for unpredictable incidents e.g. accidents or breakdowns. For these reasons, an adaptive control system is preferable [12].

5.5. Decentralized traffic signal control

As we have stated in the previous section the different signal control strategy from traffic engineering side, we will discuss now the traffic signal control based on the control engineering concepts. All mentioned types above are for isolated intersections, which means that there is no information exchanged between intersections and this structure or type of control we call it Decentralized Traffic Control. So, all of the previous mentioned control strategies' are belong to this type of control structure and the difference is the number of inputs based on the detectors quantity and functions.

Decentralized control [20-28] divides the overall system into *S* subsystems, and controls the subsystems separately based only on the local model and the information of the corresponding subsystem [11]. By dividing the original system into subsystems and by designing decentralized controllers, the full information of the whole system is also separated into parts. The information interactions between subsystems are cut off and incoming traffic from other intersections are considered constants or estimated, which results in fully isolated systems.

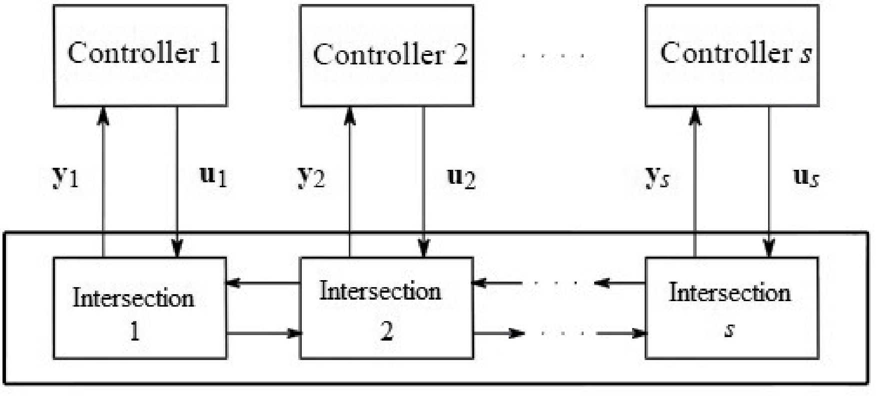


Figure 21. Decentralized Control Structure

Because the estimates of the input traffic flows from other subnetworks may be far from the real values, the local controllers may not be able to find the real optimal solutions for the subnetworks. Moreover, since the subnetworks are completely disconnected, the overall performance of the whole network will be deteriorated. As mentioned, that this type of control will do well for the isolated intersections but consider the case of an arterial network that we want to let the flow move smoothly from one intersection to the next without stoppage or minimum stoppage which of course will minimize the delay and avoid stop/start behaviour of vehicles that lead to queue accumulation.

This open another issue which we call it **Traffic signal coordination** that is normally implemented to improve the level of service of a road or a network of roads, where the spacing of signals is such that isolated signal operation would cause excessive delays, stops and loss of capacity. The popular concept is that coordinating traffic signals is simply to provide green-wave progression whereby a motorist travelling along a road receives successive green signals. While this is one of the aims, the principal purpose of coordination is to minimize overall delay and/or number of stops. This can be achieved using fixed-timing plans or using adaptive technology.

The three main components of coordinated timings are:

- Cycle time the time to complete all phases in a timing plan (a phase is any period in a cycle where non-conflicting traffic movements may run.
- Stage splits - the amount of time allocated to a phase in a cycle
- Offsets - green signals at adjacent intersections are set to occur at a given time, relative to that at a reference intersection. It depends on the distance between signals, the progression speed along the road between the signals and the queues of vehicles waiting at red signals.

In Figure.22 the left part, you can see that the vehicle is delayed at the second intersection due to an uncoordinated signal time offset while in the right figure the vehicle is not delayed at the second intersection due to a coordinated signal time offset

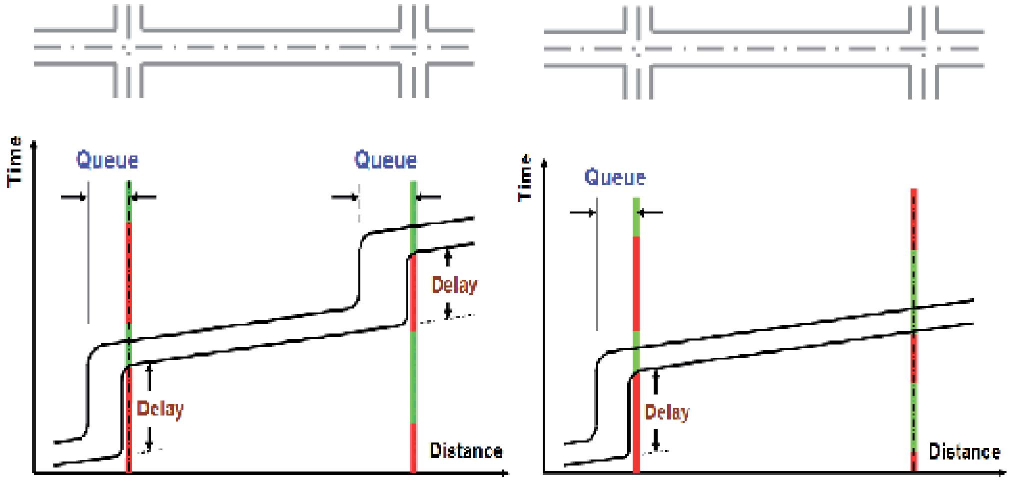


Figure 22. Two Traffic Intersection, the Left figure is without Coordination, the right figure is with coordination

Without coordination, there are frequent stops and unnecessary delays, for example:

- Vehicles pass through a green light at one intersection only to be stopped by a red light at the next intersection, causing inconsistent travel;
- Vehicles must wait through more than one green signal at an intersection due to blockages ahead;
- Vehicles must stop at a red light when there are no other vehicles or pedestrians at the cross street.
- In addition to extending travel time these situations also increase fuel consumption and emissions, as a stationary vehicle is much less efficient than a vehicle in motion.

The type and the amount of coordinated information will introduce the terms **Quasi Decentralized & Distributed Control structures**. In the Quasi Decentralized we allow minimum necessary amount of information for the purpose of coordination between intersections. For example, the current executed phase in the progressive intersection signal can be sent to the successive intersection to allow the scenario shown in Figure 22, the right part, and hence it very important to have good sensors and communication network. The distributed structure will allow more data transfer and also will required more sensors and detectors. This can be grateful from the first impression but it may add also an overhead communication which may delay the control signal or perform the control command before the arrival of the information.

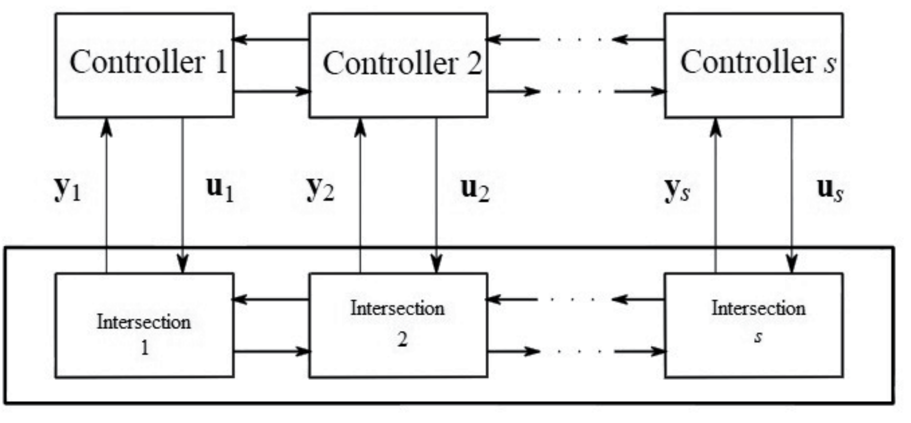


Figure 23. Distributed Control Structure

One more way is divide the large urban network into sub areas and assign controllers for each intersection, and another master or coordinator controller for that area to coordinate between local controllers. This is what we call the Hierarchical Control structure where the controllers mainly aiming at solving specified different tasks locally while the upper-level controllers will coordinate (or supervise) the subsystems from a global point of view or global objective function.

Since we are talking about the sensors network we should highlight the network link impacts on the sensors function that caused by the environment conditions or from the link itself such as the packets delay, dropout and sampling time selection. In each case there several methods to handle it and what are the controller actions in such scenarios to continue the operations of traffic signals smoothly as much as possible.

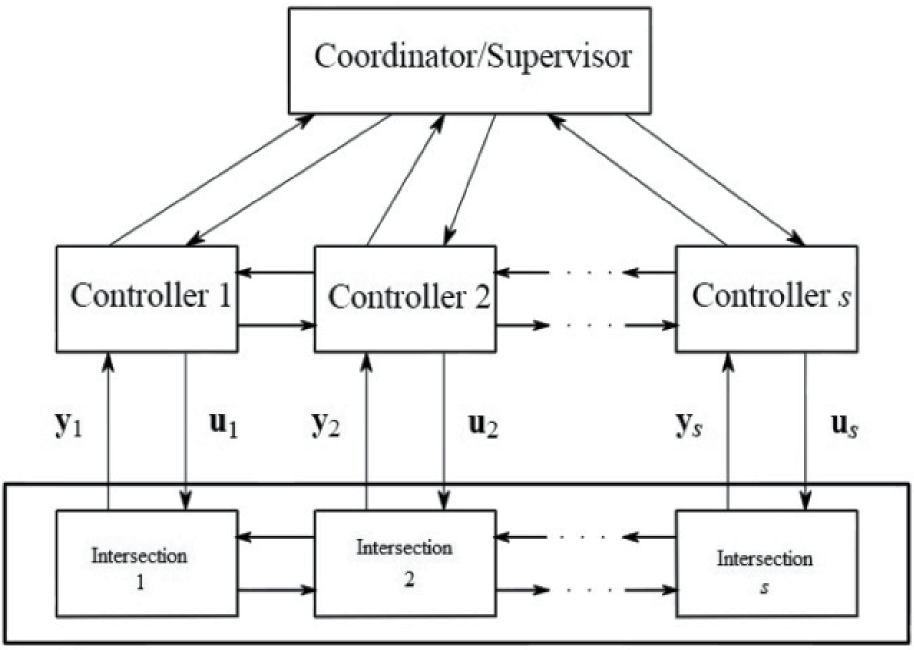


Figure 24. Hierarchical Control Structure

6. Results & discussions

In the simulation we considered 5 intersections as shown in Figure 26 and we tried to compare the results from the proposed approaches. The simulation was done using MATLAB 2008 with the following assumptions:

- Distance between each intersection is known (1 km).
- Average speed is 70 Km/h.
- Estimated time to travel from one intersection to another with 70 Km/h is around 45 sec.
- Flow of traffic is smooth and no major interruption.
- Each intersection operates in 4 phase's mode as default, which means that every two parallel directions will run at the same time to maximize the flow.
- The average of arrivals for each parallel direction will be taken as input. Figure 11 shows the considered phases ({N, S},{E,W},{NW, SE},{ES,WN}).
- Simulation runs for 30 minutes.

The arrival rate per min, the phase selection and service time (Green period T_g) are shown in the Table 1 while the level of information exchanged is shown in Table 2. In our simulation we considered the first two approaches. When we look to Table 1, we can see that at intersection 1, we started with phase 1, then by the time the flow will reach to intersection 2, which is around 45 seconds, the incoming flow plus the existing flow will move together without stoppage and same will happen at intersection 3, this explanation is shown clearly in Figure 25. That shows the beauty of Quasi Decentralized approach over the Decentralized itself, where in the Quasi we have benefited from the limited communication over a network to smooth and maximize the flow in certain direction between intersections. However, in case we are building our system on lossy communication links, then there is a chance to have a packet dropout or delay or some induced errors, still the system can take care of that but this is part is not included in this chapter.

I1			I2			I3			I4			I5		
Phase	Tg	Q	Phase	Tg	Q	Phase	Tg	Q	Phase	Tg	Q	Phase	Tg	Q
1	31	38	2	43	29	3	19	43	4	20	22	1	42	34
2	19	27	1	45	34	2	43	20	2	21	22	2	35	28
4	15	10	3	17	9	1	45	7	3	20	8	4	15	7
3	15	10	4	16	8	4	17	5	1	21	8	3	15	8

Table 1. Car Arrivals Rate /Min (Q) , Phase Selection and Service Time (T_g) for Each Intersection

	Decentralized	Quasi Decentralized	Distributed	Hierarchical
Traffic Arrival	Y	Y	Y	Y
Phase Selection (prev. intersection)	N	Y	Y	Y
Traffic Arrival previous intersection	N	N	Y	Y
Green Time	N	N	Y	Y
Traffic Jam Info	N	N	N	Y
Avg arrivals speed	N	N	Y	Y

Table 2. Data Exchange in Each Approach

I1		I2		I3		I4		I5	
Dec	Quasi	Dec	Quasi	Dec	Quasi	Dec	Quasi	Dec	Quasi
1	1	1	4	1	3	1	1	2	2
2	2	2	1	2	4	2	2	1	1
3	3	4	2	3	1	3	3	3	4
4	4	3	3	4	2	4	4	4	3

Table 3. Decentralized & Quasi Phase Selection

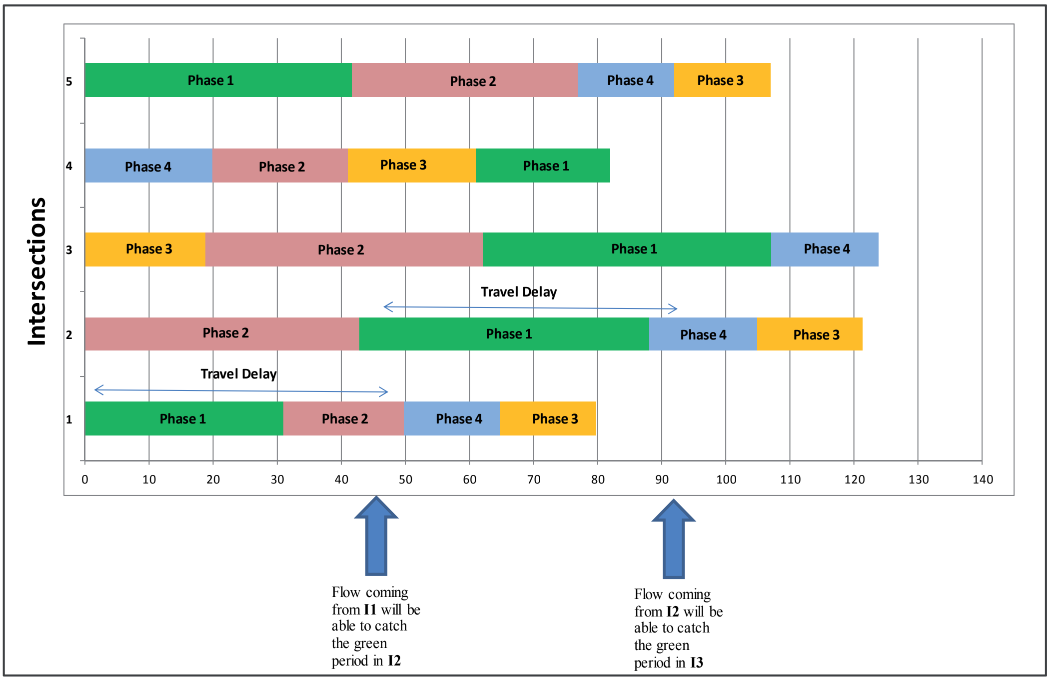


Figure 25. Phase Seclection in Each Intersection to maximaize the flow from intersection 1 up to 3

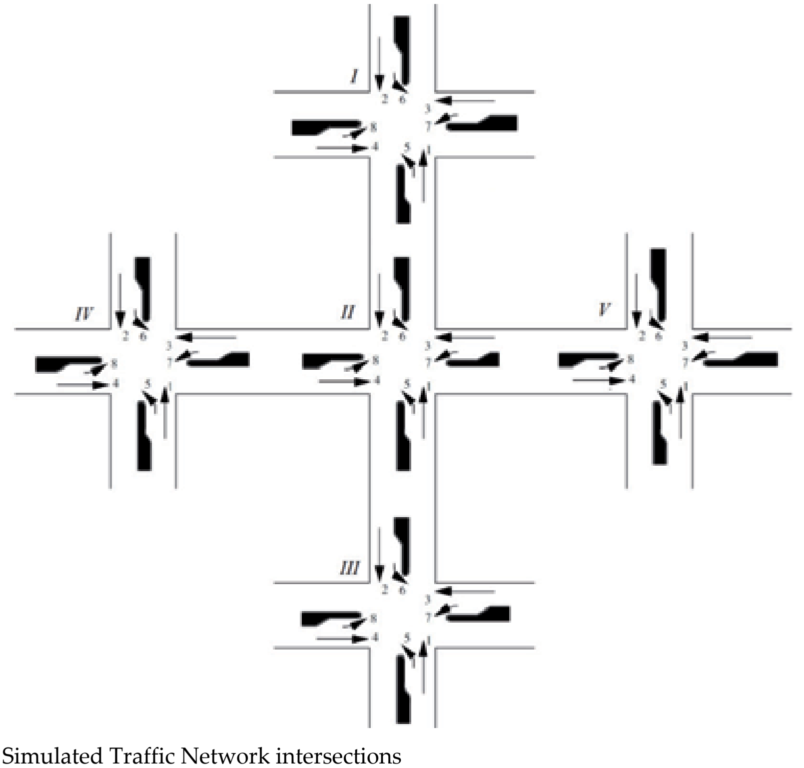


Figure 26. Simulated Traffic Network intersections

The proposed work also can work for the arterial traffic network where you series of intersections along highway and you need to maximize the flow in that heavy traffic highway with minimum number of stops. This scenario shown in Figure 28 and flow can be controlled similar to what we have done in the five intersections example.

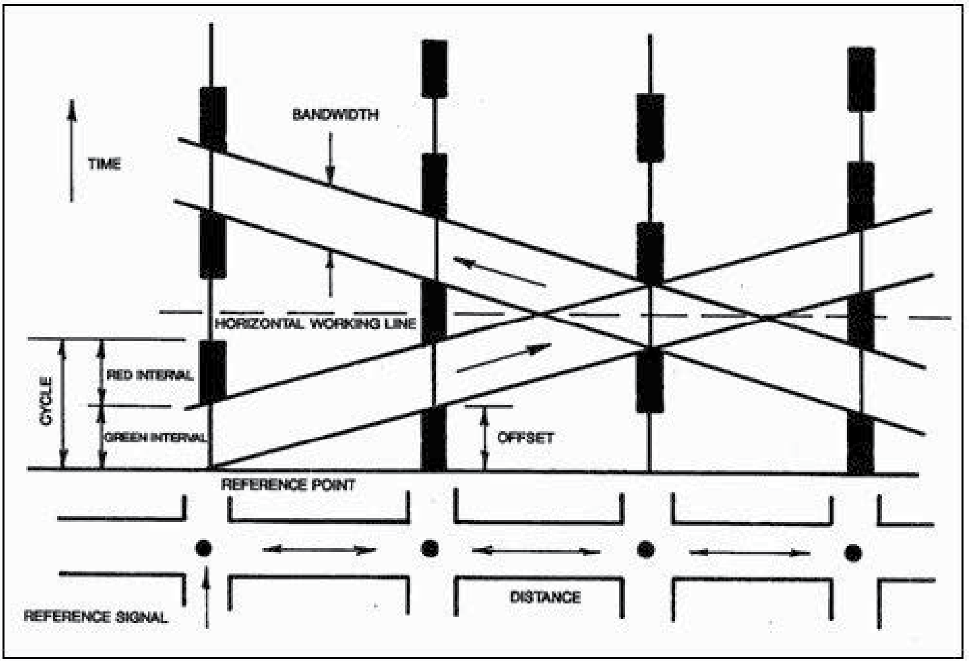


Figure 27. Arterial Traffic Network

7. Future research

In the coming work, simulation will cover more parameters and we will add the other approaches (distributed and hierarchal). Also, more discussions about the effect of network size and the media used for communication between detectors and controllers will be included. The work can be extended easily for arterial traffic network as shown in Figure 29 where also we can apply the hierarchical approach by dividing the network to zones and each zone controller will report to the master coordinator that will monitor the traffic flow in these zones.

Also we can add something about the dilemma zone as shown in figure 30 for safety purpose. A dilemma zone [29] is a range, in which a vehicle approaching the intersection during the yellow phase can neither safely clears the intersection, nor stop comfortably at the stop-line. One of the main contributors to signal-related accidents is the existence of a dilemma zone at signalized intersections. Note that both the length and the location of a dilemma zone may vary with the speed of the approaching vehicles, driver reaction times, and vehicle acceleration/deceleration rates.

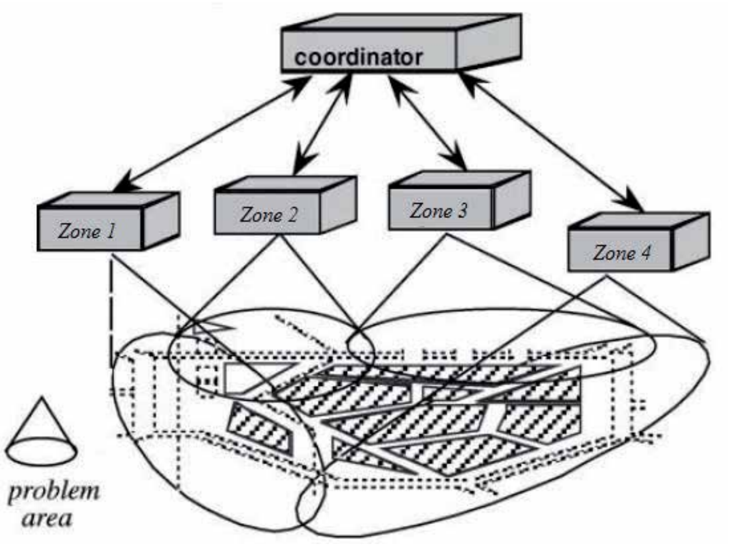


Figure 28. Hierarchal Approach

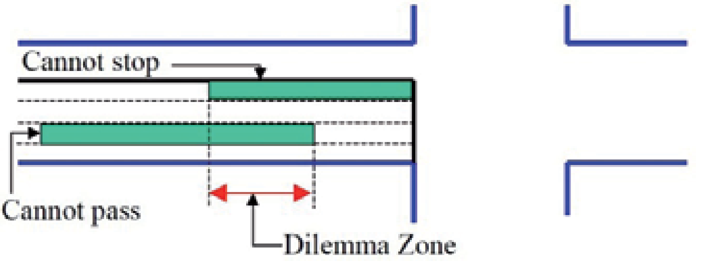


Figure 29. Dilemma Zone

8. Conclusion

In this chapter, we have covered the area of wireless sensors network and we tried to discuss in some details one important application, signalized traffic control, which depends strongly on these devices. Several types of detectors were listed with some information about each type. Then, we introduced some control strategies and concepts from both traffic and control engineering point of view. Finally, simple simulation was done to support the theoretical discussions.

9. Nomenclature

	Definition
Vehicle Presence	Presence (or absence) of a vehicle at a point on the roadway
Flow Rate	Number of vehicles passing a point on the roadway during a specified time period
Occupancy	Percent of time that a point on the roadway is occupied by a vehicle

	Definition
Speed	Distance traveled by a vehicle per unit time
Density	Number of vehicles per lane mi (km)
Jam density	Refers to extreme traffic density associated with completely stopped traffic flow, usually in the range of 185–250 vehicles per mile per lane.
Headway	Time spacing between front of successive vehicles, usually in one lane of a roadway
Queue Length	Number of vehicles stopped in a lane behind the stopline at a traffic signal
Control delay	The component of delay that results when a control signal causes a lane group to reduce sped or to stop; it is measured by comparison with the uncontrolled condition
Cycle	A complete sequence of signal indications
Cycle length	The time required for one complete sequence of signal intervals (phases).
Interval	A period of time in which all traffic signal indications remain unchanged
Lost time	The time during which an intersection is not used effectively by any movement; it is the sum of clearance lost time plus start-up lost time
Phase	The part of the signal cycle allocated to any combination of traffic movements receiving the right-of-way simultaneously during one or more intervals
Split	The percentage of a cycle length allocated to each of the various phases in a signal cycle.
Offset	The time relationship, expressed in seconds or percent of cycle length, determined by the difference between a defined point in the coordinated green and a system reference point.
Red Duration	The period in the signal cycle during which, for a given phase or lane group, the signal is red
Saturation flow rate	The equivalent hourly rate at which previously queued vehicles can traverse an intersection approach under prevailing conditions, assuming that the green signal is available at all times and no lost times are experienced
Start-up Delay	The additional time consumed by the first few vehicles in a queue at a signalized intersection above and beyond the saturation headway, because of the need to react to the initiation of the green phase and to accelerate
Extension of green duration	The amount of the change and clearance interval, at the end of the phase for a lane group, that is usable for movement of its vehicles
Green Duration	The duration of the green indication for a given movement at a signalized intersection
Change and clearance interval	The yellow plus all-red interval that occurs between phases of a traffic signal to provide for clearance of the intersection before conflicting movements are released
Clearance lost time	The time between signal phases during which an intersection is not used by any traffic

Author details

Faisal Ahmed Al-Nasser and Magdi S. Mahmoud
Systems Engineering Department, King Fahd University of Petroleum & Minerals, KFUPM, Saudi Arabia

Acknowledgement

This work is supported by the deanship of scientific research (DSR) at KFUPM through research group project No. RG-1105-1.

10. References

[1] Chong C. , Kumar S., "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247– 1256, 2003.
[2] Prasad R., Hamid S. , Michael H. , Hesham H. , Jon Y. "A New MAC Approach for Mobile Wireless Sensor Networks". 14th IST Mobile & Wireless Communication Summit, pp. 405 – 410, Dresden 19 -23 June 2005.
[3] "NEMA TS2 Traffic Controller Assemblies with NTCIP Requirements." National Electrical Manufacturers Association, 1998.
[4] Skabardonis, A., R.L. Bertini, and B.R. Gallagher. "Development and Application of Control Strategies for Signalized Intersections in Coordinated Systems." Transportation Research Record 1634, pp. 110-117. Transportation Research Board, Washington, DC, 1998.
[5] Zong T. and Thomas U., "Green Extension and Traffic Detection Schemes at Signalized Intersections". Transportation Research Record: Journal of the Transportation Research Board, No. 1978, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 16–24.
[6] Christopher M., Hiromal P., Thomas M., James R., Darcy M. ," Operational Evaluation of Wireless Magnetometer Vehicle Detectors at Signalized Intersection". Transportation Research Record: Journal of the Transportation Research Board, No. 2192, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 11–23.
[7] Dan M., Hassan C., Ryan L. , " Alternative Vehicle Detection Technologies For Traffic Signal Systems: Technical Report". Report No. FHWA/TX-09/0-5845-1, Feb. 2009.
[8] "Highway Design Manual," New York State Department of Transportation, June, 2003.
[9] L.H. Immers, S. Logghe. Traffic Flow Theory. KATHOLIEKE UNIVERSITEIT LEUVEN, Course H 111, May 2002.
[10] Tom V., Krishna R. Fundamental Relations of Traffic Flow. Introduction to Transportation Engineering, CHAPTER 31. NPTEL May 24, 2006.
[11] S. Yksel and T. Basar, "Decentralized Multi-Controller Stabilizability over Communication Channels", Proc. European Control Conference, 2007.

[12] Tubaishat, M, Shang Y., Shi H., "Adaptive Traffic Light Control with Wireless Sensor Networks" , 4th Consumer Communications and Networking Conference - CCNC, 2007.

[13] Albagul A., Hrairi M., Hidayathullah M.;"Design and Development of Sensor Based Traffic Light System", American Journal of Applied Sciences 3: 1745-1749, 2006.

[14] WEN Y, PAN J. L. , LE J. F; "Survey On Application Of Wireless Sensor Networks For Traffic Monitoring", International Conference on Transportation Engineering - ICTE, 2007.

[15] L. K. Chien, "A Real-time Dynamic Traffic. Control System Based on Wireless Sensor Network". International Conference on Parallel Processing Workshops - ICPPW, 2005.

[16] Leena S., Sudhanshu T., Himakshi A., "Time Optimization for Traffic Signal Control Using Genetic Algorithm", International Journal of Recent Trends in Engineering, Vol 2, No. 2, November 2009.

[17] Visit H., Jaturapith K., Chris H., "A Multi- Agent Approach for Intelligent Traffic-Light Control", Proc. of the World Congress on Engineering - WCE, 2007 Vol I, July , London, U.K.

[18] Faisal A., Hosam R., "Simulation of Dynamic Traffic Control System Based on Wireless Sensor Network", IEEE Symposium on Computers & Informatics 2011 (ISCI 2011), pp. 302 – 307.

[19] Magdi S. Mahmoud, "Decentralized Stabilization of Interconnected Nonlinear Systems with Time-Varying Delays", IEEE Transactions Automatic Control, Vol. 54, No. 11, November Issue, pp. 2663-2668.

[20] Magdi S. Mahmoud, "Decentralized Reliable Control of Interconnected Systems with Time-Varying Delays", Journal of Optimization Theory and Applications, Vol. 143, No. 3, December Issue, pp. 497-518.

[21] Magdi S. Mahmoud and Naif B Almutairi, "Decentralized Stabilization of Interconnected Systems with Time-Varying Delays", European J. Control, Vol. 15, No. 6, December Issue, pp. 624-633.

[22] Magdi S. Mahmoud, "Decentralized Reliable Control of Interconnected Systems with Time-Varying Delays", Journal of Optimization Theory and Applications, Vol. 143, No. 3, December Issue, pp. 497-518.

[23] Ali Y. Al-Rayyah and Magdi S. Mahmoud, "Decentralized Reliable Control of Interconnected Time-Delays Systems against Sensor Failures", Journal of Optimization Theory and Applications, Vol. 147, No. 2, November Issue, pp. 318-336.

[24] Magdi S. Mahmoud, "Decentralized Output-Feedback Stabilization for Interconnected Discrete Systems with Unknown Delays", Optimal Control Applications and Methods, Vol. 31, No. 6, November-December Issue, pp. 529-545.

[25] Magdi S. Mahmoud, (Decentralized Systems with Design Constraints), Springer-Verlag, UK, May 2011. ISBN 978-0-85729-289-6

[26] Magdi S. Mahmoud, "Design of Control Strategies for Robust Dynamic Routing in Traffic Networks", IET Control Theory and Applications, Vol. 5, No. 15, September Issue, pp. 1716-1728.

[27] Bo Liu, Yuanqing Xia, Magdi S Mahmoud and Harris Wu, "New Predictive Control Scheme for Networked Control Systems", Journal of Circuits, Systems and Signal Processing, DOI: 10.1007/s00034-011-9359-9.
[28] Magdi S. Mahmoud and F. M. AL-Sunni, "Global Stability Results of Discrete Recurrent Neural Networks with Interval Delay", IMA J. Mathematical Control and Information.
[29] Liu C., Tao H., Tabacek. "Empirical observations of dynamic dilemma zones at signalized intersections". The 86th meeting of the Transportation Research Board, July 2006.

Edited by Mohammad Matin

The aim of this book is to present few important issues of WSNs, from the application, design and technology points of view. The book highlights power efficient design issues related to wireless sensor networks, the existing WSN applications, and discusses the research efforts being undertaken in this field which put the reader in good pace to be able to understand more advanced research and make a contribution in this field for themselves. It is believed that this book serves as a comprehensive reference for graduate and undergraduate senior students who seek to learn latest development in wireless sensor networks.

Photo by chombosan / iStock

IntechOpen

ISBN 978-953-51-5573-7