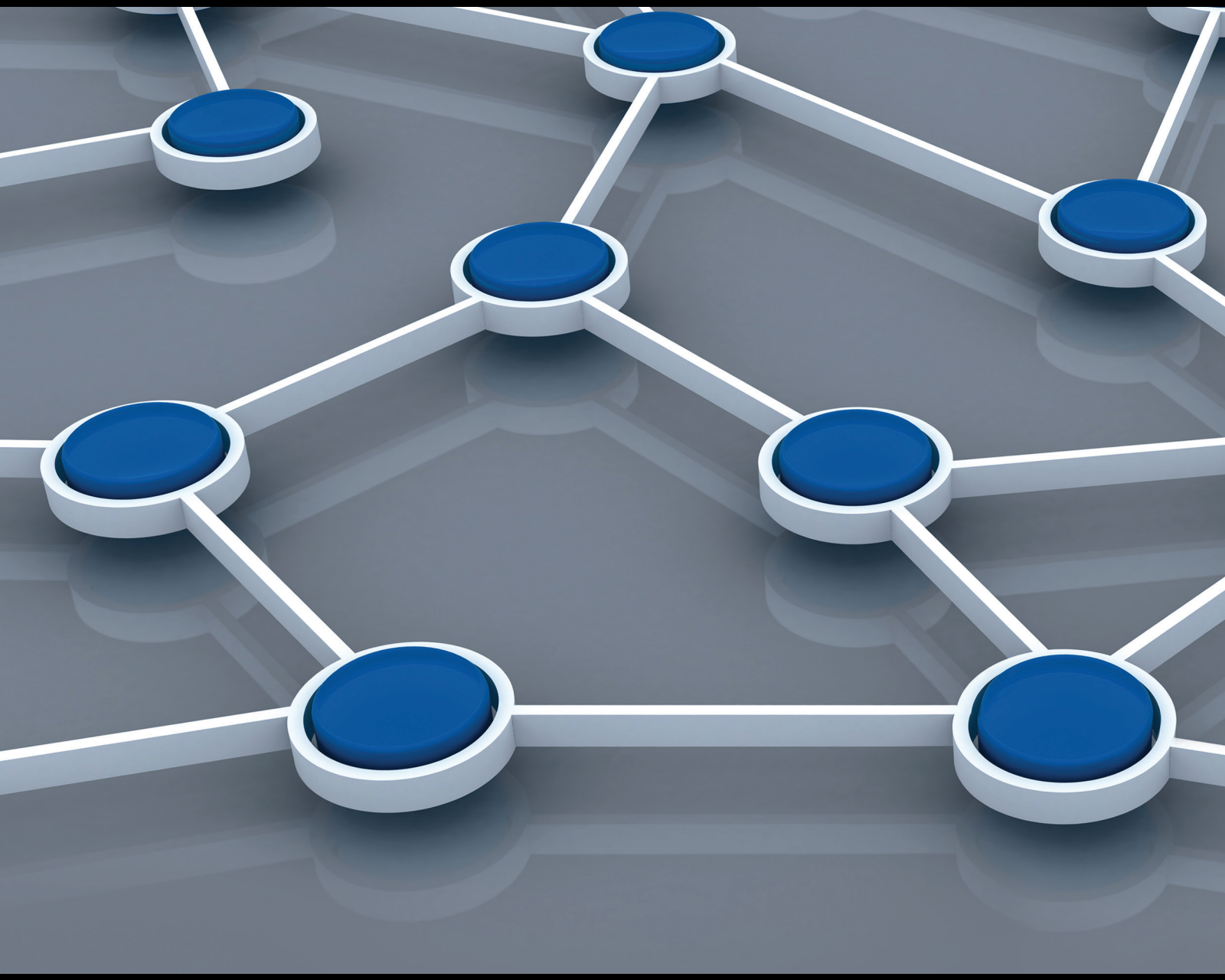
International Journal of Distributed Sensor Networks

Protocols and Architectures for Next-Generation Wireless Sensor Networks

Guest Editors: Sana Ullah, Joel J. P. C. Rodrigues, Farrukh Aslam Khan, Christos Verikouki, and Zuqing Zhu



Protocols and Architectures for Next-Generation Wireless Sensor Networks

International Journal of Distributed Sensor Networks

Protocols and Architectures for Next-Generation Wireless Sensor Networks

Guest Editors: Sana Ullah, Joel J. P. C. Rodrigues, Farrukh Aslam Khan, Christos Verikouki, and Zuqing Zhu

Editorial Board

Contents

Protocols and Architectures for Next-Generation Wireless Sensor Networks, Sana Ullah, Joel J. P. C. Rodrigues, Farrukh Aslam Khan, Christos Verikouki, and Zuqing Zhu
Volume 2014, Article ID 705470, 3 pages

Load Balanced Routing for Lifetime Maximization in Mobile Wireless Sensor Networks, Saifullah Khalid, Ashraf Masood, Faisal Bashir Hussain, Haider Abbas, and Abdul Ghafoor
Volume 2014, Article ID 979086, 12 pages

Wireless M-Bus Sensor Networks for Smart Water Grids: Analysis and Results, S. Spinsante, S. Squartini, L. Gabrielli, M. Pizzichini, E. Gambi, and F. Piazza
Volume 2014, Article ID 579271, 16 pages

ZEQoS: A New Energy and QoS-Aware Routing Protocol for Communication of Sensor Devices in Healthcare System, Zahoor Ali Khan, Shyamala Sivakumar, William Phillips, and Bill Robertson
Volume 2014, Article ID 627689, 18 pages

Energy Consumption Optimisation for Duty-Cycled Schemes in Shadowed Environments, Tatjana Predojev, Jesus Alonso-Zarate, Mischa Dohler, and Luis Alonso
Volume 2014, Article ID 709135, 10 pages

Game-Theoretic Based Distributed Scheduling Algorithms for Minimum Coverage Breach in Directional Sensor Networks, Jin Li, Kun Yue, Weiyi Liu, and Qing Liu
Volume 2014, Article ID 341309, 12 pages

Altruistic Backoff: Collision Avoidance for Receiver-Initiated MAC Protocols for Wireless Sensor Networks, Xenofon Fafoutis, Charalampos Orfanidis, and Nicola Dragoni
Volume 2014, Article ID 576401, 11 pages

A QoS-Based Wireless Multimedia Sensor Cluster Protocol, Juan R. Diaz, Jaime Lloret, Jose M. Jimenez, and Joel J. P. C. Rodrigues
Volume 2014, Article ID 480372, 17 pages

WSN4QoL: A WSN-Oriented Healthcare System Architecture, S. Tennina, M. Di Renzo, E. Kartsakli, F. Graziosi, A. S. Lalos, A. Antonopoulos, P. V. Mekikis, and L. Alonso
Volume 2014, Article ID 503417, 16 pages

An Improved User Authentication Protocol for Healthcare Services via Wireless Medical Sensor Networks, Muhammad Khurram Khan and Saru Kumari
Volume 2014, Article ID 347169, 10 pages

Wireless HDLC Protocol for Energy-Efficient Large-Scale Linear Wireless Sensor Networks, Daniel Mihai Toma, Joaquin del Rio, and Antoni Mànuel
Volume 2014, Article ID 916073, 14 pages

Protocol and Architecture to Bring Things into Internet of Things, Ángel Asensio, Álvaro Marco, Rubén Blasco, and Roberto Casas
Volume 2014, Article ID 158252, 18 pages

Energy-Efficient Node Selection Algorithms with Correlation Optimization in Wireless Sensor Networks, Hongju Cheng, Zhihuang Su, Daqiang Zhang, Jaime Lloret, and Zhiyong Yu
Volume 2014, Article ID 576573, 14 pages

Securing Cognitive Wireless Sensor Networks: A Survey, Alexandros Fragkiadakis, Vangelis Angelakis, and Elias Z. Tragos
Volume 2014, Article ID 393248, 12 pages

Collection Tree Extension of Reactive Routing Protocol for Low-Power and Lossy Networks, Jiazi Yi and Thomas Clausen
Volume 2014, Article ID 352421, 12 pages

Design and Experiment Analysis of a Hadoop-Based Video Transcoding System for Next-Generation Wireless Sensor Networks, Haoyu Xu, Liangyou Wang, and Huang Xie
Volume 2014, Article ID 151564, 7 pages

RFID Localization Using Angle of Arrival Cluster Forming, Waleed Alsalih, Abdallah Alma'aitah, and Wadha Alkhater
Volume 2014, Article ID 269596, 8 pages

Energy Efficient and Load Balanced Routing for Wireless Multihop Network Applications, Vahid Nazari Talooki, Jonathan Rodriguez, and Hugo Marques
Volume 2014, Article ID 927659, 13 pages

A Cross-Layer Approach to Minimize the Energy Consumption in Wireless Sensor Networks, Luca Catarinucci, Riccardo Colella, Giuseppe Del Fiore, Luca Mainetti, Vincenzo Mighali, Luigi Patrono, and Maria Laura Stefanizzi
Volume 2014, Article ID 268284, 11 pages

A Survey on Deployment Algorithms in Underwater Acoustic Sensor Networks, Guangjie Han, Chenyu Zhang, Lei Shu, Ning Sun, and Qingwu Li
Volume 2013, Article ID 314049, 11 pages

Maximizing Network Lifetime of Directional Sensor Networks Considering Coverage Reliability, Joon-Min Gil, Jong Hyuk Park, and Young-Sik Jeong
Volume 2013, Article ID 583753, 8 pages

A QoS Model for a RFID Enabled Application with Next-Generation Sensors for Manufacturing Systems, Anna Kang, Jong Hyuk Park, Leonard Barolli, and Hwa-Young Jeong
Volume 2013, Article ID 829691, 6 pages

Editorial

Protocols and Architectures for Next-Generation Wireless Sensor Networks

Sana Ullah,[1] Joel J. P. C. Rodrigues,[2,3] Farrukh Aslam Khan,[4] Christos Verikoukis,[5] and Zuqing Zhu[6]

[1] *CISTER Research Unit, ISEP/IPP, Rua Dr. Antonio Bernardino de Almeida 431, 4200-072 Porto, Portugal*
[2] *Instituto de Telecomunicações, University of Beira Interior, 6201-001 Covilhã, Portugal*
[3] *University ITMO, Saint-Petersburg 197101, Russia*
[4] *National University of Computer and Emerging Sciences, A. K. Brohi Road H-11/4, Islamabad 44000, Pakistan*
[5] *Telecommunications Technological Centre of Catalonia, Avenida Carl Friedrich Gauss 7, 08860-Castelldefels, Spain*
[6] *Department of EEIS, University of Science and Technology of China, Hefei, Anhui 230027, China*

Correspondence should be addressed to Sana Ullah; sanajcs@hotmail.com

Received 10 August 2014; Accepted 10 August 2014; Published 23 December 2014

The massive deployment of Wireless Sensor Networks (WSNs) is expected to increase exponentially in the next few years, allowing millions of wireless devices to work autonomously for new applications. These next-generation WSNs are also expected to interact with other devices such as RFID tags, home appliances, cars, and mobile equipment. When integrated with the cloud, these networks could provide pervasive and ubiquitous services to the users by providing powerful and unlimited storage infrastructure. The current challenges for next-generation WSNs are scalability, decentralization, resource scarcity, heterogeneity, and dynamicity. These requirements and challenges cannot be fulfilled by traditional WSNs. In addition, the network, Medium Access Control (MAC), and Physical (PHY) layer protocols developed for traditional WSNs are not applicable to the next-generation WSNs, where millions of battery-powered nodes should operate for longer durations. In other words, the next-generation WSNs require the development of novel protocols at each layer that must be able to extend the network lifetime from months to years. These protocols may allow seamless integration of next-generation WSNs with other networks and platforms including internet of things and cloud computing.

The aim of this special issue was to attract high quality papers on protocols and architectures for the next-generation WSNs. We have received forty one articles, which were rigorously peer-reviewed by experts, and have finally selected twenty one articles for publication. The paper entitled *"Load Balanced Routing for Lifetime Maximization in Mobile Wireless Sensor Networks"* proposes a novel lifetime maximization protocol for heterogeneous and homogenous networks with uncontrolled mobility by considering residual energy, traffic load, and mobility of nodes. Simulation results show that the proposed scheme provides significant improvement in network lifetime, data packet latency, and load balancing compared to minimum hop routing and greedy forwarding schemes. The paper entitled "*Wireless M-Bus Sensor Networks for Smart Water Grids: Analysis and Results*" investigates wireless metering bus protocol for consideration in future smart water grids. Simulation and experimental results show the effectiveness and feasibility of the proposed protocol. The paper entitled *"ZEQoS: A New Energy and QoS-Aware Routing Protocol for Communication of Sensor Devices in Healthcare System*" proposes a novel and QoS-aware routing protocol using two main modules and three algorithms for resource allocation. Simulations conducted in a real hospital scenario using Castalia 3.2 show that the proposed protocol offers good performance in terms of throughput and packet dropping rate at MAC and network layers. The paper entitled *"Energy Consumption Optimisation for Duty-Cycled Schemes in Shadowed Environments"* proposes a metric for low-power wireless links by considering shadow fading and truncated

ARQ schemes. NS-3 simulations are used to determine the optimal operating regions for direct, multihop, and CDC-ARQ forwarding. The paper entitled *"Game-Theoretic Based Distributed Scheduling Algorithms for Minimum Coverage Breach in Directional Sensor Networks"* formulates the problem of direction set K-Cover as a direction scheduling game and proposes synchronous and asynchronous game-theoretic based distributed algorithms. Experimental results show that the proposed algorithms (Nash equilibria) provide a near-optimal and well-balanced solution. The paper entitled *"Altruistic Backoff: Collision Avoidance for Receiver-Initiated MAC Protocols for Wireless Sensor Networks"* presents a novel collision avoidance method called altruistic backoff that reduces a significant amount of energy by minimizing the idle listening period of the nodes. The performance of the proposed backoff method is validated by several experiments using TI eZ430-rf2500 nodes. The paper entitled *"A QoS-Based Wireless Multimedia Sensor Cluster Protocol"* presents a new protocol for delivering multimedia stream features and guaranteeing quality of communication. Real experiments show performance of the proposed protocol for several video and audio cases in terms of bandwidth, delay, and jitter.

The paper entitled "*WSN4QoL: A WSN-Oriented Healthcare System Architecture*" presents network coding and distributed localization solutions for achieving efficiency in communication and indoor people tracking. Preliminary results show the efficiency of the proposed solutions. The paper entitled *"An Improved User Authentication Protocol for Healthcare Services via Wireless Medical Sensor Networks"* proposes an improved user authentication scheme for healthcare applications and proves that the proposed scheme eliminates security problems that are identified in the existing security schemes. The paper entitled "*Wireless HDLC Protocol for Energy-Efficient Large-Scale Linear Wireless Sensor Networks*" proposes a wireless HDLC architecture that supports half-duplex communication and point-to-point and multipoint networking. A hardware prototype for self-powered wireless sensors, based on XBee PRO modules, is developed for a large-scale infrastructure monitoring system. The paper entitled *"Protocol and Architecture to Bring Things into Internet of Things"* proposes a Communication Things Protocol (CTP) that provides interoperability among things with different communication standards. CTP considers an ontological representation and interaction model of things and implementation feasibility in standard communication protocols. Performance analysis shows the feasibility of CTP in terms of energetic cost, data efficiency, and message latency. The paper entitled *"Energy-Efficient Node Selection Algorithms with Correlation Optimization in Wireless Sensor Networks"* first proposes a new cover set balance algorithm to select a set of active nodes with partially ordered tuple, and then it proposes a new algorithm to find the correlated node set for a given node. In addition, a high residual energy first algorithm is proposed for reducing the number of active nodes. Experiments show that the proposed algorithms significantly increase the network lifetime. The paper entitled *"Securing Cognitive Wireless Sensor Networks: A Survey*" presents an overview of the recent progress in the area of cognitive sensor networks and highlights open research issues and challenges. The paper entitled *"Collection Tree Extension of Reactive Routing Protocol for Low-Power and Lossy Networks"* presents an extension to the Lightweight On-Demand Ad hoc Distance Vector Routing Protocol-Next Generation (LOADng) routing protocol for efficient construction of a collection tree. The extended LOADng imposes minimal overhead and complexity and avoids complications of unidirectional links in the collection tree. The complexity, security, and interoperability of the proposed protocol are analyzed using extensive simulations. The paper entitled *"Design and Experiment Analysis of a Hadoop-Based Video Transcoding System for Next-Generation Wireless Sensor Networks"* presents a hadoop-based video transcoding system for accommodating hundreds of high-definition video streams in the next-generation sensor networks. Experimental results show that there is an optimal value of the number of mappers, which is closely related to the file size. In addition, it is also shown that the time consumption of video transcoding depends on the duration of video files rather than on their sizes.

The paper entitled *"RFID Localization Using Angle of Arrival Cluster Forming"* presents a test-bed comparison of power control and RSSI distance estimation approaches for active RFID tags. It also presents an angle of arrival cluster forming localization approach that utilizes the angle of arrival of the tag's signal and the reader's transmission power control in order to localize the active tags. The paper entitled *"Energy Efficient and Load Balanced Routing for Wireless Multihop Network Applications*" presents a novel, energy-efficient, and traffic balancing routing protocol that provides a weighted and flexible trade-off between energy consumption and load dispersion. Simulation results show that the proposed protocol achieves high energy efficiency, decreases the number of failed nodes, and extends the network lifetime. The paper entitled *"A Cross-Layer Approach to Minimize the Energy Consumption in Wireless Sensor Networks"* presents a cross-layer solution to reduce idle listening period by triggering the node whenever a packet is detected. The wakeup circuit or MAC scheduler wakes up the nodes using a commercial power detector connected to the nodes. Experiments are conducted to show effectiveness of the proposed solution. The paper entitled *"A Survey on Deployment Algorithms in Underwater Acoustic Sensor Networks"* overviews the most recent advances of deployment algorithms in underwater acoustic sensor networks. It classifies the algorithms into static deployment, self-adjustment deployment, and movement-assisted deployment. The paper entitled *"Maximizing Network Lifetime of Directional Sensor Networks Considering Coverage Reliability"* addresses the Directional Cover-sets with Coverage Reliability (DCCR) problem by presenting a Coverage Reliability model and a Directional Coverage and Reliability (DCR) greedy algorithm. The Coverage Reliability model considers the detection probability of each node in the cover-sets. The DCR algorithm solves the DCCR problem. Simulation results show that the proposed approaches increase network lifetime while guaranteeing the minimum coverage reliability. The paper entitled "*A QoS Model for a RFID Enabled Application with Next-Generation Sensors for Manufacturing Systems"* proposes a quality model for RFID systems. The criterion for the quality model is borrowed

from ISO 9126 and the DeLone and McLean models. The proposed model consists of functionality, reliability, usability, efficiency, maintainability, and business criteria. When the manufacturing system addresses these criteria and satisfies users and developers, RFID system benefits can be obtained.

Acknowledgments

The authors would like to thank all the reviewers who have rigorously reviewed the papers and helped them select the best papers for publication. They would also like to thank the editorial staff for their continuous support.

Sana Ullah
Joel J. P. C. Rodrigues
Farrukh Aslam Khan
Christos Verikoukis
Zuqing Zhu

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 979086, 12 pages
http://dx.doi.org/10.1155/2014/979086

Research Article

Load Balanced Routing for Lifetime Maximization in Mobile Wireless Sensor Networks

Saifullah Khalid,[1] Ashraf Masood,[1] Faisal Bashir Hussain,[2] Haider Abbas,[1,3] and Abdul Ghafoor[1]

[1] *National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan*
[2] *Department of Computer Science, Bahria University, Islamabad 44000, Pakistan*
[3] *Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia*

Correspondence should be addressed to Saifullah Khalid; sukhalid@gmail.com

Received 1 November 2013; Revised 13 February 2014; Accepted 9 April 2014; Published 15 July 2014

Academic Editor: Farrukh Aslam Khan

Challenge of efficient protocol design for energy constrained wireless sensor networks is addressed through application specific cross-layer designs. This design approach along with strong design assumptions limits application of protocols in universal scenarios and affects their practicality. With proliferation of embedded mobile sensors in consumer devices, a changed application paradigm requires generic protocols capable of managing greater device heterogeneousness and mobility. In this paper, we propose a novel lifetime maximization protocol for mobile sensor networks with uncontrolled mobility considering residual energy, traffic load, and mobility of a node. The protocol being generic is equally applicable to heterogeneous, homogenous, static, and mobile sensor networks. It can handle event driven as well as continuous traffic flow applications. Simulation results show that proposed scheme outperforms minimum hop routing and greedy forwarding in terms of network lifetime, data packet latency, and load balance while maintaining comparable throughput.

1. Introduction

Wireless sensor networks (WSN) have wide range of applications in many areas of daily life [1]. Routing protocols for WSN are generally application specific and cross-layer design approach is adopted to achieve efficiency. Application specific cross-layer protocols have strong design assumptions and are not suitable for universal scenarios. This improved performance comes at the cost of design modularity, stability, and robustness. Cross-layer design involves complex interactions among multiple network layers ranging from physical to application layer. Suitable models to describe these interactions are still being investigated. Unless these models are available, cross-layer architecture would find little acceptance in universal context. On the contrary, in layered architecture, complex problems are easily solved by breaking into simple ones. Layered architecture leverages modular, loosely coupled adaptable designs and has secured deeper acceptance in industry.

WSN are traditionally considered as no or quasimobility networks. However, mobility can leverage greater benefits in terms of improved coverage with sparse sensor deployment, healing of topological defects, energy efficiency, and increased application domains. Few areas utilizing mobility are urban sensing, assisted living and residential monitoring, industrial automation, and mobile sensor based wide area monitoring. WSN mobility is characterized as controlled or uncontrolled. Controlled mobility is used for efficient data collection and healing of topological defects. Mobility is deemed to tradeoff delay to achieve energy and resource efficiency. The approach is less suitable for applications with hard realtime constraints. Uncontrolled mobility is relatively less researched but is important in context of proliferation of sensors in consumer devices, mobile phones, personal data assistants, and special purpose platforms. Use of mobile sensors with uncontrolled mobility in routing tasks is so far rather limited. These sensors can be utilized in applications like people centric urban sensing and assisted living. In urban

sensing [2] environment, data can be very speedily passed back to static infrastructure using sensors embedded into user devices carried by robots or vehicles.

A sensor network with uncontrolled mobility represents a very large heterogeneous network comprising mobile as well as static networks. Static networks are connected through mobile devices carried by robots or vehicles. This heterogeneous network is connected to back haul infrastructure to form a very large cooperative network in contrast to small scale application specific sensor implementations. Because of overwhelming mobility and heterogeneity of involved devices, managing interactions among network elements is a very complex task.

In mobile wireless sensor networks, due to heavier maintenance cost, static or preconfigured routing is less suitable. Also, proactive approach is not feasible for event driven sensor networks where information generated by sensor nodes is not known a priori and depends on the arbitrary occurrence of events. In [3], authors argue that data packet sizes in WSN are smaller as opposed to other computer networks and signalling overhead in this case becomes significant compared to data traffic. On-demand protocols have less maintenance cost but generate a lot of signalling traffic for discovery of new paths. This assertion can be true in case of scaler data but is not valid for high end futuristic as well as multimedia sensors.

In this work, we propose an on-demand routing scheme for mobile sensor networks with uncontrolled mobility. The protocol considers, in its path selection, the residual energy, traffic load, and mobility of a node. Main design objective of the proposed routing scheme is to maximize network lifetime. The protocol can be applied in static as well as mobile scenarios. It keeps practical limitations in view and does not compromise efficiency. The protocol suits equally event driven as well as continuous monitoring applications. Simulation analysis shows that the proposed scheme can effectively handle sensor network mobility, increase network lifetime, and decrease data packet latency.

The rest of the paper is organized as follows: Section 2 summarizes related work; in Section 3, we present network model; Section 4 describes design of proposed routing scheme and its operation; performance evaluation and analysis of results are given in Section 5. Section 6 concludes the paper and highlights future work directions.

2. Related Work

In this section, we present related work that surveys issues of routing in mobile sensor networks (MWSN) and load balanced routing and energy aware routing.

Research in MWSN gained momentum in recent years especially in mobility assisted data collection and urban sensing. A comprehensive survey of routing protocols for MWSN is available in [4]. Moreover, surveys of mobility based communication techniques are available in [5–7] and mobility models can be found in [8]. In [6], requirements, merits, and demerits of three mobility based schemes are compared. A comprehensive survey of data collection techniques using mobile elements is presented in [5]. The authors categorize mobile elements as relocatable nodes used to heal topological defects, mobile data collectors for data collection, and mobile peers for sensing and routing tasks.

The authors in [9] study use of mobile relays as resource provisioning method to extend lifetime of sensor network in a large dense network. They conclude that use of one energy rich mobile relay can extend network lifetime up to four times of that of static network. The work in [10–13] investigates lifetime maximization problem using mobile sink; especially, issues related to finding optimum sink route or trajectory are addressed. These proposals utilize controlled mobility for efficient data collection and do not take into account nodes embedded into mobile platforms having uncontrolled mobility.

Developing a large scale general purpose sensor network in urban setting for the general public is studied in [2]. Authors propose network architecture based on opportunistic sensor network paradigm capable of supporting urban sensing with widespread people centric applications and heterogeneity in devices. Sensor speed, direction of move, and location are used to select a sensor for delegation or tasking.

The authors in [14] propose a mobility aware routing protocol where mobility is used to form sink cluster and during route discovery process. A cluster based routing protocol for a low mobility homogenous sensor network is presented in [15]. The nodes are considered to follow random mobility model. Zone head is elected based on mobility factor which is taken as ratio of zone changes to position changes within a zone. The scheme considers node speed and location information for determining mobility factor. In our routing scheme, mobility factor is one of the factors considered to select the next hop node. However, our technique of mobility factor determination considers node speed and does not require node location information exchange. The scheme [15] tries to balance energy consumption by considering the number of times a node has acted as zone head whereas in our scheme balance is achieved by considering the traffic load a node receives.

In [16], authors have studied the problem of reducing energy consumption by flow augmentation to balance energy utilization across network. This scheme uses residual energy of nodes as basic admission control criteria. Selection of nodes on the said criteria can balance energy consumption but results in longer source to destination paths and increases latency of information delivery. One of the earliest proposals on energy aware routing [17] considers clustered network topology and utilizes topological information for this purpose. Performance of such protocols is severely affected by mobility as topology constantly changes resulting in significant topology maintenance overhead.

Load balance and local congestion control is investigated in [17]. It considers two identical metrics, that is, maximum connections per relay and overall relay load. These metrics help to increase lifetime of relay node and avert packet loss by avoiding overcommitted nodes from becoming relay. But limiting maximum connections per relay node can result in coverage issues across the network. E-WLBR [18] is a proactive routing protocol, in which load balance is

achieved by distributing traffic among the next hop neighbors according to their load handling capacity determined in terms of residual energy levels. Each node notifies its load handling capacity during initialization phase. Protocol in its present form is less suitable for a network with mobile nodes and bears disadvantage of proactive protocols for scalable networks. Authors in [3] show that sending traffic on multiple paths can reduce significant energy consumption. Candidate paths for forwarding traffic are determined based on multiple weighted factors. However, existence of completely disjoint multiple paths can only enhance performance but it is totally dependent on network topology. Also, in mobile networks, using multiple paths will increase route maintenance overhead as mobility can affect all paths. In another scheme [19], one or two next hop neighbors are selected according to hybrid routing metric and traffic is then distributed among these selected neighbors in round robin or weighted round robin manner. Round robin achieves per packet load balancing whereas in weighted round robin traffic is distributed according to assigned weights. In mobile networks, candidate neighboring nodes can change frequently and maintaining even two hop nodes information can result in enhanced energy overhead. LEAR [20] considers a number of active routes through a relay node for load balancing and routes multimedia traffic on fully or partially disjoint paths. However, LEAR does not consider realistic traffic load on a relay node but assumes that each flow is identical, having same data rate. Reference [21] surveys load balanced routing strategies and highlights that this issue still requires significant research.

3. Network Model

In this section, network model is presented and highlights the assumptions and terminologies used in the proposed routing scheme. Moreover, techniques of utilizing and estimating node mobility, energy, and load are described.

The target network is of heterogeneous nature consisting of mix of high and low end sensors. The high end sensors possess relatively better processing and energy resources whereas low end sensors are constrained in these resources. The network nodes are assumed to be deployed according to flat or random topology as depicted in Figure 3. In target network, the majority of network nodes are static while the remaining are mobile. The nodes have inherent mobility detection mechanism in place. Mobility is not used for resource provisioning or data collection; rather, the sensors are onboard a mobile platform, for example, sensing robot, vehicle, consumer device, or an aerial platform.

The terminologies used in this work are defined as follows.

(i) *Static Sensor Network*. It is a wireless sensor network where all nodes are static.

(ii) *Mobile Node*. It is a sensor node embedded in a sensing robot, a vehicle, a consumer device, or an aerial platform. The node not only carries out sensing tasks but also relays messages from other nodes.

(iii) *Mobility Detection*. The node is capable of detecting mobility and for this purpose it either has GPS or relative position detection mechanism in place.

(iv) *Low Mobility Network*. It is a network which consists of majority of static and some mobile sensor nodes. The mobile nodes may follow random or group mobility models.

(v) *Medium Mobility Network*. It is a network which consists of equal number of mobile and static sensor nodes. Mobile nodes follow random or group mobility models.

3.1. Sensor Network Mobility. Besides benefits, mobility also poses challenges in protocol design as it affects route stability and route maintenance cost. For efficient protocol design, mobility must be taken into account to avoid establishing routes through mobile nodes, thus conserving energy in frequent maintenance. Node mobility is characterized in terms of mobility factor and is estimated based on location information using approaches discussed below. These schemes have varying degree of computation complexity, accuracy, and need for information exchange. For WSN, a lesser complex scheme requiring no additional information exchange is a better choice. However, accuracy may be improved by taking moving average of mobility measure over certain period of time. Mobility prediction approaches are as follows.

3.1.1. Transitions Count. The approach assumes that sensor network is divided into zones which may be defined according to a specific criterion. Node mobility is measured in terms of number of transitions of a mobile node across different zones. The scheme has limitations in case of group motion where although the nodes move across different zones they may still maintain association or link with their neighbors. This approach also requires location information exchange.

3.1.2. Remoteness. To capture the notion of relative mobility, the concept of remoteness is introduced in [22]. Here the mobility factor is determined in terms of rate of link change. If nodes in a zone are in group motion, average link change is minimal. The node movement in such scenarios does not affect association of node with a zone or a link. So the remoteness of a node from its neighbors can be treated as a measure of mobility and is given as

$$S = \frac{1}{N}\sum_{i=1}^{N} d_i(t) \quad \forall N > 1, \tag{1}$$

where $d_i(t) = (1/n)\sum_{j=1}^{n} d_{ij}(t)$ for all $n \neq 0$. Let $d_{ij}(t)$ be the distance at time t of node i from jth neighbor; $d_i(t)$ is average distance of a node from its n neighbors. By considering average distance over N time intervals link change rate can be determined. The approach requires exchange of location information among nodes.

3.1.3. Speed. Node speed may also be used as measure of node mobility. However, such a representation has limitations especially in case of group motion. If the nodes are in group motion at constant speed, they do not break link despite motion. In other cases, a node itself may be static with the least mobility factor but its neighbors may move out breaking the link. However, this approach does not require any location information exchange and can be very easily calculated:

$$S = \frac{1}{N}\sum_{i=1}^{N} v_i(t) \quad \forall N > 1, \tag{2}$$

where $v_i(t) = (1/T)\sum_{t=1}^{T} d_i(t)$ for all $T \neq 0$ represent sensor node velocity over an interval T and N is the number of old samples being considered.

3.2. Energy Aware and Lifetime Maximization Based Routing. WSN have extreme constraints of energy; therefore, energy efficiency is the main design objective during protocol design. Several approaches for energy consumption are in practice [16, 23, 24]. One of the well-known metrics for decreasing energy consumption and latency is the selection of minimum hop paths. However, this results in premature death of those nodes that are frequently used in minimum hop routing paths. On the contrary, energy balanced algorithms utilize suboptimal paths to maximize network lifetime [16, 23].

3.3. Load Balanced Routing. Network load balance is another important factor for lifetime maximization [3, 17, 19, 20]. Load aware routing helps to conserve energy by avoiding collisions and overcome delays caused by local congestion. Load balance is achieved by spreading traffic either on multiple paths or by avoiding overcommitted nodes as relays. Multipath routing has related overheads and energy costs, whereas later approach is simple and can be implemented without global knowledge. This metric helps in achieving longer network lifetime by avoiding overloaded nodes to participate in routing. The traffic load of a node is given as follows:

$$l_o = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_i(t), \tag{3}$$

where $f_i(t)$ is the counting function over time interval T and N is the number of old samples being considered.

4. Proposed Routing Scheme

The scheme uses hybrid cost function for routing decisions. The hybrid metric is formed based on the factors discussed in Section 3. Summarized description of these factors is as follows.

(1) Mobility Factor. Competing approaches to estimate mobility have been discussed in Section 3.1. Considering low mobility, energy expenditure for location information exchange and lesser computation cost mobility factor based on node speed are used in this scheme. Node mobility S is given as below:

$$S = \frac{1}{N}\sum_{i=1}^{N} v_i(t) \quad \forall N > 1, \tag{4}$$

where $v_i(t) = (1/T)\sum_{t=1}^{T} d_i(t)$ for all $T \neq 0$ represent sensor node velocity over an interval T and N is the number of old samples being considered. Windowed exponential moving average of node speed helps smoothing transients over a period of time T. A node with the least mobility is considered a better candidate for the next hop. This helps increasing link lifetime and adds to longer network lifetime by saving energy required for frequent maintenance of broken routes.

(2) Residual Energy. Energy aware routing and network lifetime have been discussed in Section 3.2. This scheme considers residual energy level of a node for selecting it as the next hop. Initially, once energy level is high, this factor has little role to play in routing decisions. However, as energy depletes, it becomes a dominating consideration. This metric allows minimum hop routing initially and thus improved network delays. In our case, a node with greater residual energy is a preferred choice as the next hop node.

(3) Node Load Figure. By considering state of load being handled by a node, energy wastage due to collisions and delay in servicing packets can be reduced. Load balancing helps in achieving better network lifetime and is discussed in Section 3.3. The load at a node is given as follows:

$$l_o = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_i(t), \tag{5}$$

where $f_i(t)$ is the counting function over time interval T and N is the number of old samples being considered. Windowed moving average of load helps smoothing transients over a period of time T. A node with the least load is preferred to be selected as the next hop node. This helps in increasing network lifetime.

(4) Path Length Constraint. Proposed scheme allows use of suboptimal paths in favour of balanced energy consumption and longer network lifetime. However, a likely pitfall of forming extremely nonoptimal paths is prevented by using path length constraint. A minimum cost routing path is selected only if it is within X hops of the shortest path; otherwise, the shortest path routing is performed. X is the maximum allowed hop deviation from the minimum or shortest path between source and destination. It is dependent on network size and node density. In [20], authors have shown based on experimental results that a deviation of up to four hops from the shortest path can give better throughput in an average size network if the shortest path is not suitable due to high traffic load.

4.1. Cost Function. Hybrid routing metric based on factors discussed here and also in Section 2 is as follows:

$$R_m = w_m\frac{S}{S_{\max}} + w_o\frac{l_o}{R_{\max}} + \frac{w_r}{e_r} \quad \forall e_r \neq 0. \tag{6}$$

Require: Values of R_m and Hop Count (HopCnt)
Ensure: Select a path with maximum residual energy, least mobility and least congestion
  if Source and RREQ ID not in routing table **then**
    Setup reverse path with source
    Broadcast packet to neighbors
  else
    if R_m < Previous R_m & HopCnt < min HopCnt + X (X is the deviation from the shortest path) **then**
      Setup reverse path with source of duplicate RREQ
      Drop packet
    else
      Drop packet
    end if
  end if

Algorithm 1: Route discovery process.

R_m is hybrid routing metric used for routing decisions in proposed scheme, S is mobility factor, l_o is the traffic load, and e_r is residual energy of a node. w_m, w_o , and w_r are weights for mobility factor, traffic load, and residual energy, respectively. These weights can be selected according to type of network and to alter the contribution of a particular factor in overall decision making. For example, in a high mobility network, in order to increase network lifetime, w_m can be made comparatively bigger than w_o and w_r. Similarly, if energy constraints are sever, then w_r might be made bigger.

$S_{\max}$ and $R_{\max}$ are maximum node speed and application reporting rate. These factors are used to normalize node speed and load. The value of $S_{\max}$ can be estimated for a particular application. However, maximum traffic handled by a node in mobile multihop network is quite difficult to estimate especially because of forwarding load component. Therefore, normalized load factor may not lie in 0 to 1 interval and result in biased routing decisions. In order to overcome such a situation, dynamic adjustment of weights based on either fuzzy logic or analytical hierarchical process (AHP) may be used. However, it would entail additional processing cost. In other cases, these weights can be experimentally selected for a particular application.

Each node in the network calculates its routing metric R_m based on its residual energy, load, and mobility factor. This figure is updated after short intervals of time. Selection of metric update interval has effect on selection of optimal route as well as network lifetime. Each node shares its metric R_m with other nodes by packing it in route request (RREQ) and route reply (RREP) messages. Node with the least R_m is considered a better next hop node. Initially, once the residual energy is high, routing decisions are based mainly on sensor load and mobility factor. Once the energy is depleted, residual energy factor becomes significant; thus it ensures that nodes with lesser energy are not selected for relaying packets.

4.2. Route Discovery. Once a node has to send data or it receives RREQ, for another node for which it does not have an active route, it broadcasts RREQ message to its neighbors. Besides other information, it inserts value of R_m in RREQ packet. Based on metric value received in RREQ, a node decides to select the source node as the next hop or otherwise. The route discovery scheme is given in Algorithm 1 and is also depicted in Figure 1.

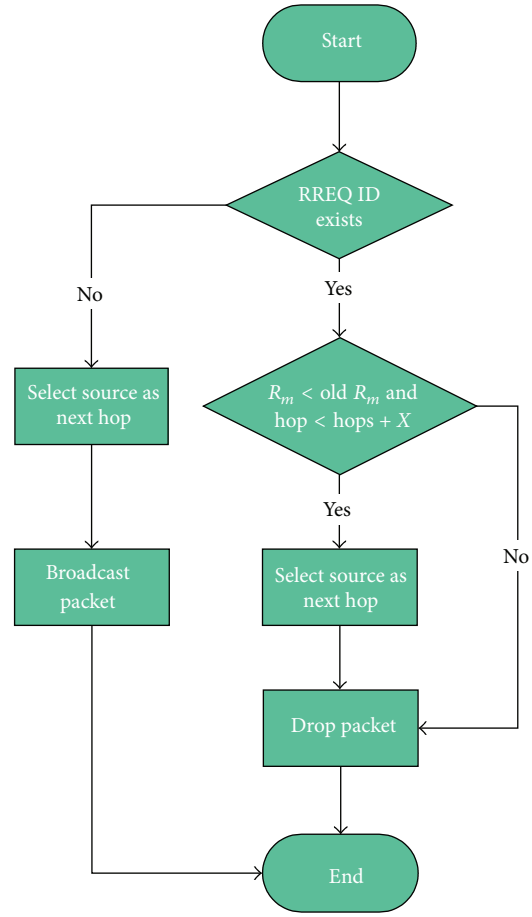


Figure 1: Route discovery process.

4.3. Operation. Load balanced routing (LBR) protocol is an on-demand routing protocol which is designed to maximize network lifetime for sensor networks with mobile elements.

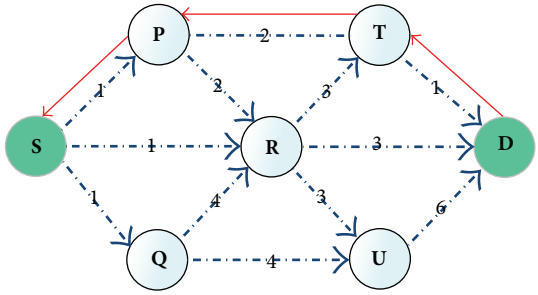


Figure 2: Dotted edges show forward dissemination of RREQ from source **S** to sink **D** whereas solid edges (red colored) represent the data route established. The number on the edge shows value of routing metric R_m. Instead of the shortest path through node **R**, a suboptimal path along nodes **T** and **P** is established.

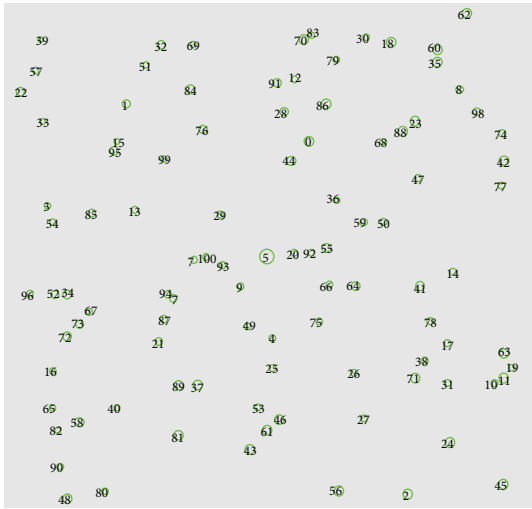


Figure 3: Sensing field: showing 100 randomly deployed sensors in 200 × 200 m area. Node 5 in bigger circle represents sink. Five such deployments are used in this evaluation. The sink and the nodes were randomly deployed in each case. The simulation results presented in Section 5 are the average of the results obtained for each of these scenarios.

Routing decisions are made based on hybrid routing factor R_m given in (6). A node with the least R_m is preferred as the next hop node.

When a node has data to transmit, it broadcasts RREQ message to its neighbors. At a neighbor node, three cases are possible. In the first case, a neighbor node may be the destination itself, so it sends route reply (RREP) message and records value of R_m beside other essential information (source and destination addresses, hop count, RREQ ID and sequence number, etc.) in its routing table. In the second case, a node may not have received the RREQ message previously; then it broadcasts RREQ to its neighbors and records essential information in its routing table. In the third case, where the node has already received RREQ, it drops it; however, if the old value of metric is bigger than the newly received, it updates value of R_m in its routing table and also sets up backward pointer to this node. If the hop count received in

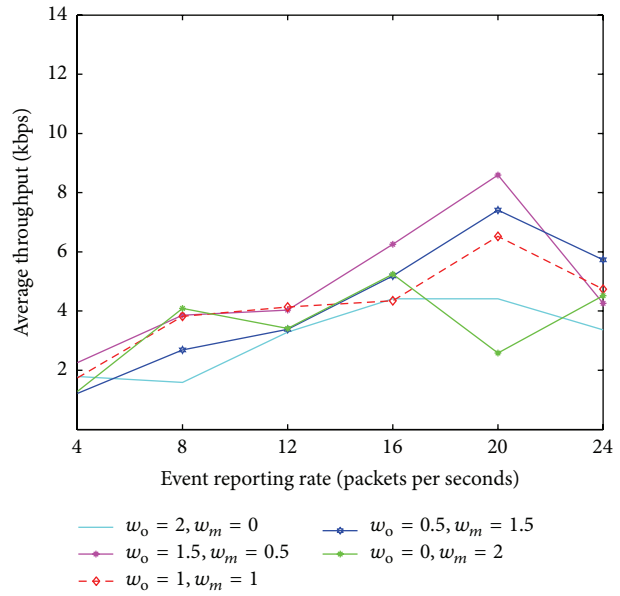


Figure 4: Performance with different weights.

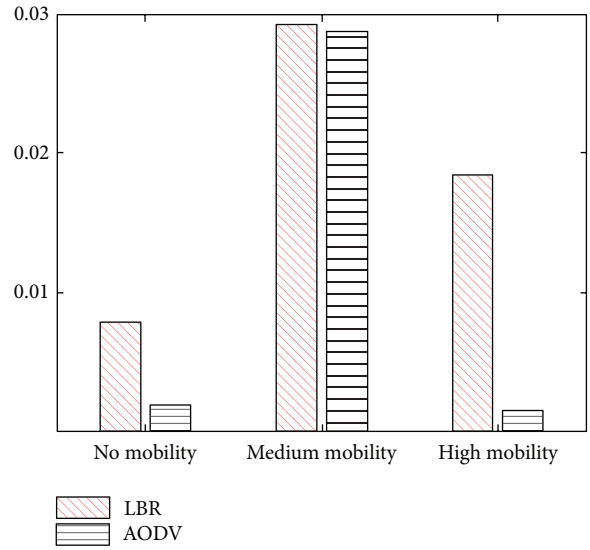
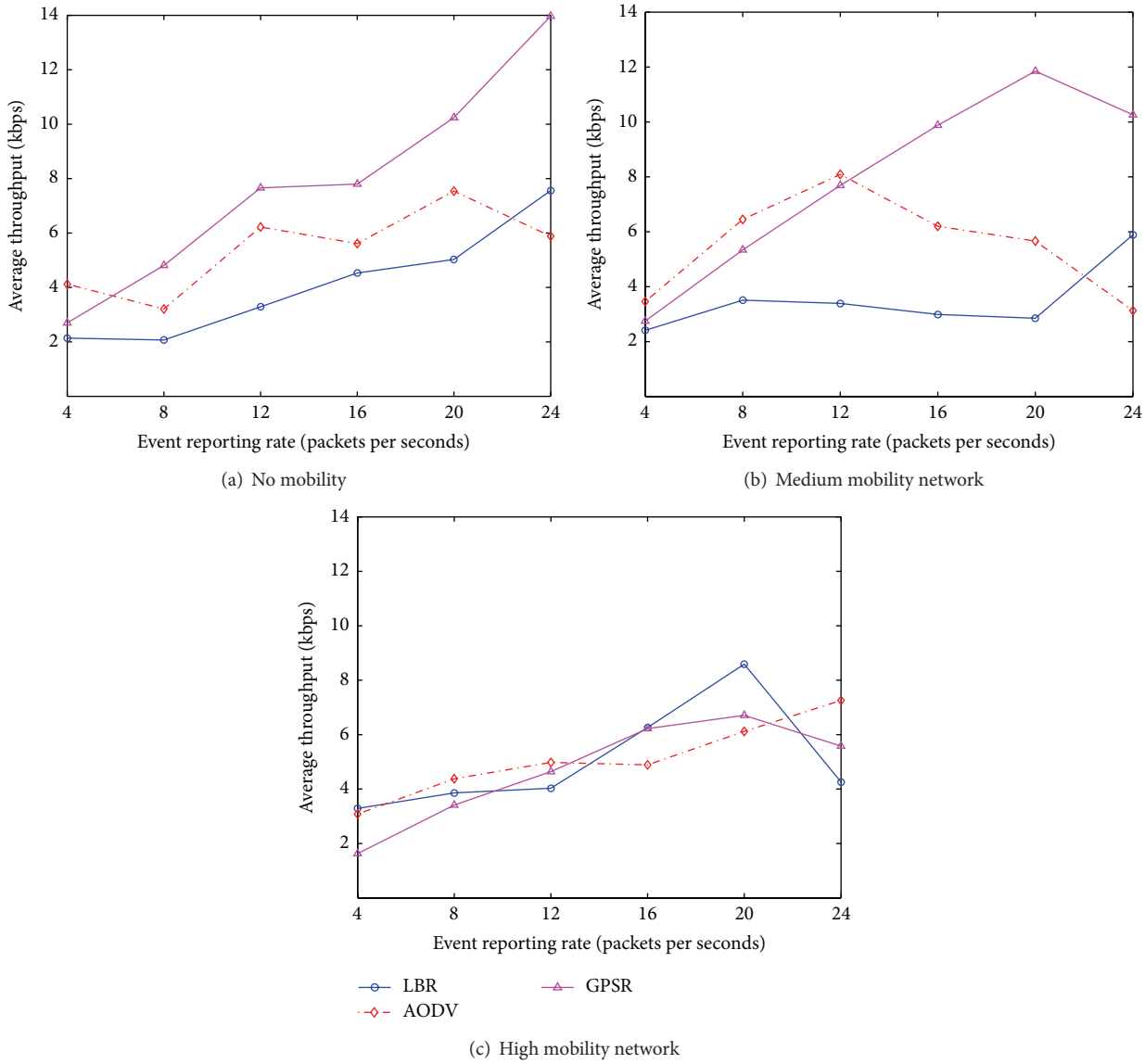


Figure 5: Packet delivery success rate.

RREQ is more than X hops of the minimum hop path, then backward pointer is not reset even if value of R_m is lesser than previously set path. So the least mobility, highest residual energy, and least busy path from source to destination are established. The established backward route or pointer is also subject to path length constraint. The path establishment operation is depicted for one source and sink in Figure 2.

5. Performance Evaluation

In this section, we report performance of LBR compared to shortest path routing and greedy forwarding protocols. For this purpose, AODV and GPSR protocols are used. We study the effect of different weights, network size, traffic load,



(a) No mobility

(b) Medium mobility network

(c) High mobility network

Figure 6: Average data throughput.

and mobility on protocol performance. The traffic load is increased gradually by increasing application reporting rate, that is, number of packets per second. However, to capture effect of mobility, the evaluation is done in a static network, a low mobility network with 25% mobile nodes, and a high mobility network with 50% mobile nodes. The evaluation metrics, experimental setup, simulation parameters, results, and their analysis are presented in this section.

5.1. Performance Metrics. The metrics of throughput, data latency, network lifetime, and load balance are used to measure performance. These evaluation metrics are defined as follows.

5.1.1. Throughput. It is the measure of average number of bits per second of application data received at sink during the entire simulation period.

5.1.2. Latency. The end to end delay of data packets is the average time taken by data packets during flow from source to sink. It also includes the time taken during discovery and establishment of route to sink.

5.1.3. Network Lifetime. Network lifetime is determined in a number of ways including time till the death of the first node, certain percentage of nodes, or time till all of the nodes die. During this evaluation, we consider sensor deaths over time and plot the remaining alive nodes over simulation duration. Moreover, nodes with energy less than 0.001 joules are considered dead because of their inability to transmit sensed data due to low energy reserve.

5.1.4. Load Balance. The metric represents the distribution of traffic load per node or across segments of sensing field. In our study, we consider number of packets per node and

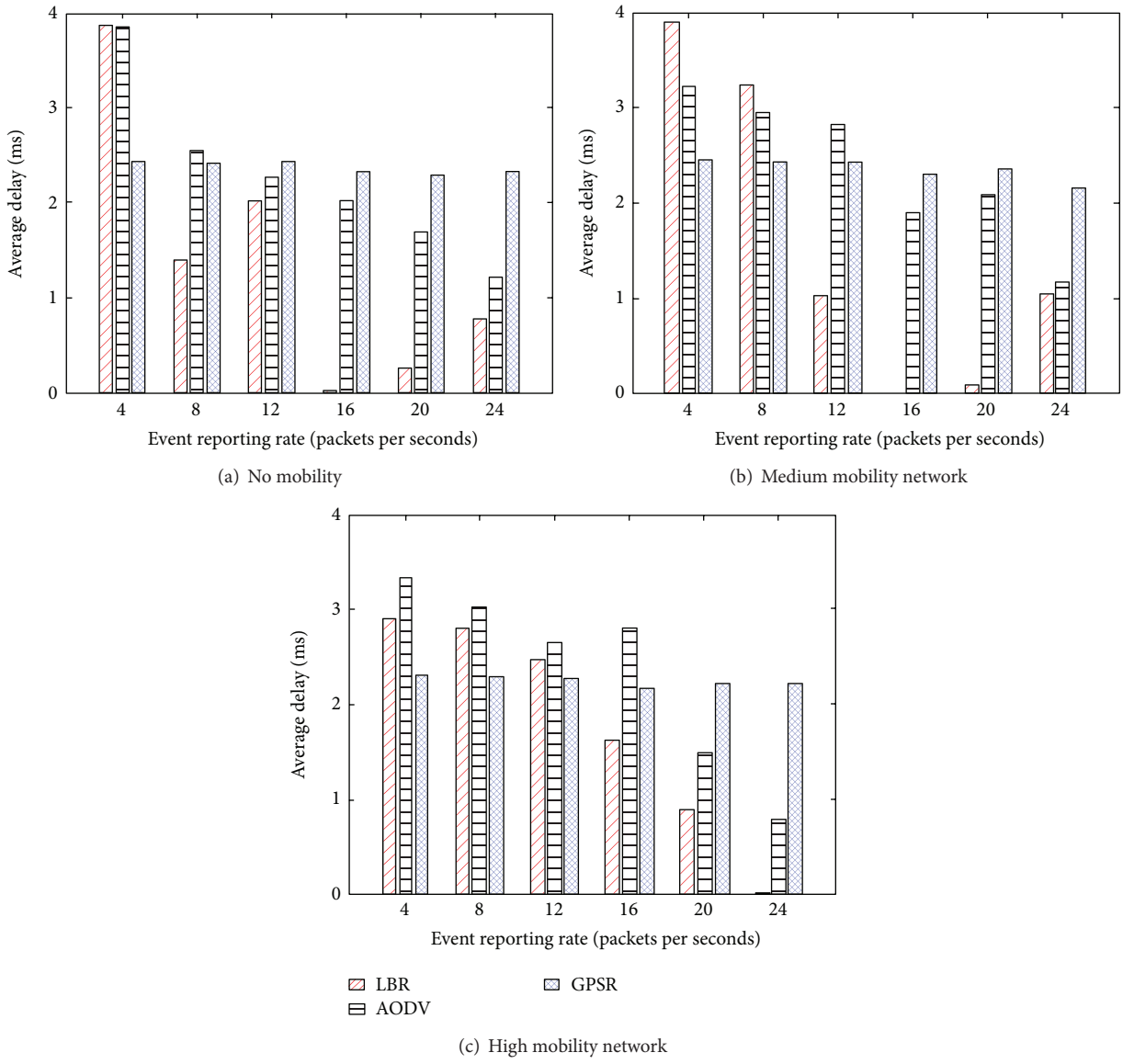


FIGURE 7: Average end to end delay of data packets: graph shows logarithm of delay.

plot per node load normalized by total number of packets successfully received at sink.

5.2. Experimental Setup. Network simulator (NS-2) is used to evaluate protocol performance. The simulations are conducted over five sensing fields of 200 by 200 meters. Each contains one sink and 99 randomly deployed nodes. The sink is considered to be static while other nodes are a mix of static and mobile nodes. Sensing field with randomly deployed nodes is shown in Figure 3. During this evaluation, one hundred random events scattered over sensing field and staggered randomly over simulation time are considered. After occurrence, the event is assumed to be reported for 60 seconds at specified rate. Summary of simulation parameters is given in Table 1 and key parameters are described as follows.

5.2.1. Energy. The experiments are conducted assuming homogenous networks and all sensors are set to have initial energy of 2 joules except sink which is assumed to have no energy limitation. The performance of proposed scheme is assumed to be even better in case of heterogeneous network due to energy aware routing.

5.2.2. Mobility. Protocol performance is evaluated in static and a network with 25 and 50 percent mobile nodes. The sink is assumed to be static although protocol imposes no such limitation and relative performance measures obtained for static sink are equally applicable to a mobile sink also. The mobile nodes are assumed to follow random way point mobility model at average speed of 1.11 m/s.

Table 1: Simulation parameters.

Parameter	Value
Topology	Flat or random
Sensing field	200 × 200 m
Simulation time	900 s
Number of sensor nodes	100
Radio communication range	20 m
Initial node energy except sink	2 j
Traffic type	CBR over UDP
Application reporting rate	4, 8, 12, 16, 20, and 24 Pkts/s
Packet size	100 bytes
Number of events	100
Event reporting time	60 s
Mobility model	Random way point
Node speed	Min 0.25, max 3, and mean 1.1 m/s
MAC type	IEEE 802.15.4
Radio propagation model	Two-ray ground reflection
T[a]	1 s
N[b]	4
X[c]	5

[a]Metric update interval.
[b]Samples for moving average.
[c]Path length constraint.

Table 2: Weights selection.

Weights[a]	Average throughput (kbps)[b]
(2, 0)	3.52
(1.5, 0.5)	4.87
(0.5, 1.5)	4.27
(1, 1)	4.22
(0, 2)	3.14

[a]Values for w_o and w_m, for example, (2, 0), represent $w_o = 2$ and $w_m = 0$. w_r was set as 0.2.
[b]Averaged over 5 scenarios and 6 application reporting rates in high mobility network.

Table 3: Performance versus network size.

Number of nodes	Average throughput (kbps)[a]
25	4.15
50	6.22
75	4.30
100	3.34

[a]Averaged over five scenarios and six application reporting rates in high mobility network.

5.2.3. Application Reporting Rate. In case of scaler traffic in WSN, the data rate is considered to range up to few kilo bits; therefor, scheme is evaluated in a scenario where event reporting rate is set to 4, 8, 12, 16, 20, and 24 packets per second.

5.3. Performance with Different Weights. In this work, the weights as in (6) are experimentally selected with an objective to increase overall network throughput. The results are taken over five high mobility network scenarios by varying application reporting rates. A number of events occurring simultaneously are 6.7 (100 events of 60-second duration, uniformly distributed over simulation time of 900 seconds, and thus $(100/900) * 60 = 6.7$ events). Average throughput for different w_o and w_m is shown in Table 2. Maximum average throughput is obtained with $w_o = 1.5$ and $w_m = 0.5$. Protocol performance with different weights is depicted in Figure 4. It can be seen that, if more weightage is given to load, for example, $w_o = 2$ and $w_m = 0$, protocol performs better for application with higher reporting rates compared to $w_o = 0$ and $w_m = 2$ which show better results for lower reporting rates as load component is not considered. The results for different weights start to decline for reporting rate above 20 packets per second because, in case of 802.15.4 MAC, application data rate in NS-2 is approximately 120 kbps [25, 26]; this limit is even less for multihop case, so, with event reporting rate of 20 packets per second, application data rate handled by sink may reach maximum capacity ($6.7 * 20 * 100 * 8 = 104.7$ kbps where 6.7 is number of events and event reporting rate is 20 packets of 100 bytes). Because of this, an unpredictable throughput spike is observed at 24 packets for $w_o = 0$ and $w_m = 2$. Overall better results are obtained with $w_o = 1.5$ and $w_m = 0.5$, so these weights are taken as reference during further simulation.

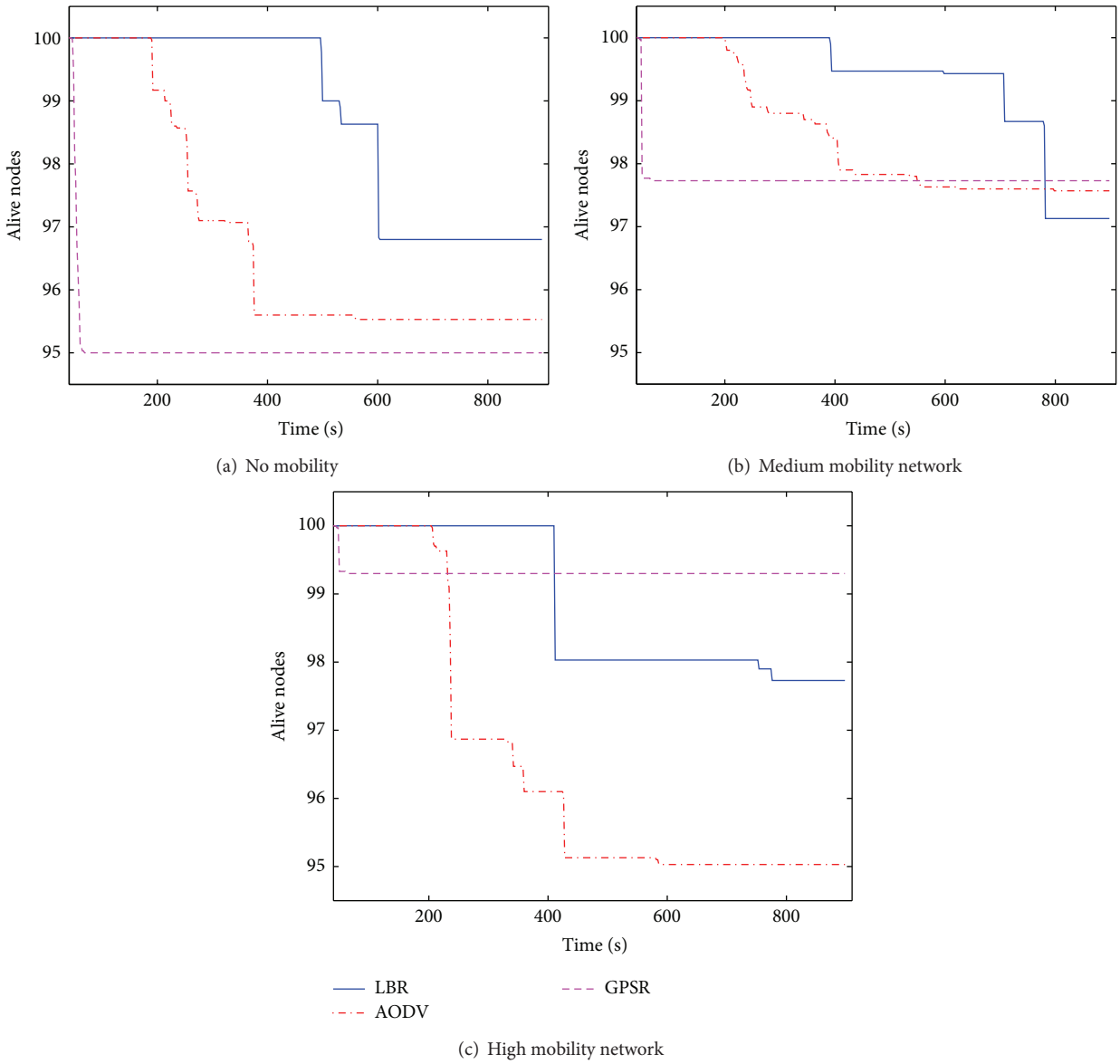
5.4. Routing Overhead and Packet Delivery Ratio. Route request message of LBR is similar to AODV except additional field for piggy packing value of hybrid routing metric R_m as per (6) which would require few additional bytes depending upon platform where LBR is implemented. However, since route request flooding is suppressed in LBR, its routing overhead is lesser compared to other reactive protocols.

For packet delivery success rate, five random deployment scenarios and hundred events reported at eight packets per second are considered. Performance under no, medium, and high mobility settings is shown in Figure 5. LBR achieves better delivery ratio for static, medium, and high mobility networks compared to AODV because of its congestion and mobility resilience.

5.5. Performance versus Network Size. For simulation, networks consisting of 25, 50, 75, and 100 nodes are taken. Each experiment is repeated five times with different random deployment scenarios. In each scenario, 50% nodes are made mobile and 10 traffic flows spanning over 500 seconds are used. Average results in each case are shown in Table 3. With increase in network nodes, the average throughput increases; however, beyond a threshold, throughput starts to decrease as contention for medium access increases resulting in packet drops.

5.6. Evaluation of Results

5.6.1. Throughput. The data throughput is shown in Figure 6. LBR has better throughput than AODV and GPSR for higher mobility and traffic loads, whereas, for lower traffic loads



(a) No mobility

(b) Medium mobility network

(c) High mobility network

Figure 8: Average number of alive nodes over time.

and mobility, it has comparable results. While the increase in traffic load congestion occurs, LBR being congestion resilient avoids overcommitted nodes, thus achieving greater throughput. Similarly, with increase in mobility, link lifetime decreases and maintenance cost is also increased but LBR can still maintain better data throughput because of mobility awareness.

5.6.2. Latency. Average end to end data latency is shown in Figure 7. LBR has better average data latency because of load balanced routing which helps to avoid overloaded nodes, improves packet service time, and thus reduces resultant delays. For lower traffic and mobility, LBR has comparable results, but, with increase in traffic load because of congestion resilience, it avoids overcommitted nodes and thus better latency figures are obtained.

5.6.3. Network Lifetime. The LBR achieves much longer network lifetime than the minimum hop routing and greedy forwarding; the results are shown in Figure 8. For both AODV and GPSR, network partitioning occurs much earlier as indicated by constant number of alive nodes. LBR performs better in terms of both first node death and number of dead nodes. Because of comparable throughput and much longer network lifetime, new scheme can transfer much more data contents compared to both other protocols before network partitioning.

5.6.4. Load Balance. The network load distribution of three protocols is shown in Figure 9. LBR achieves more even distribution of load despite greater content transfer. Owing to load aware routing, proposed scheme prevents taking of same route and distributes load across nodes. This even

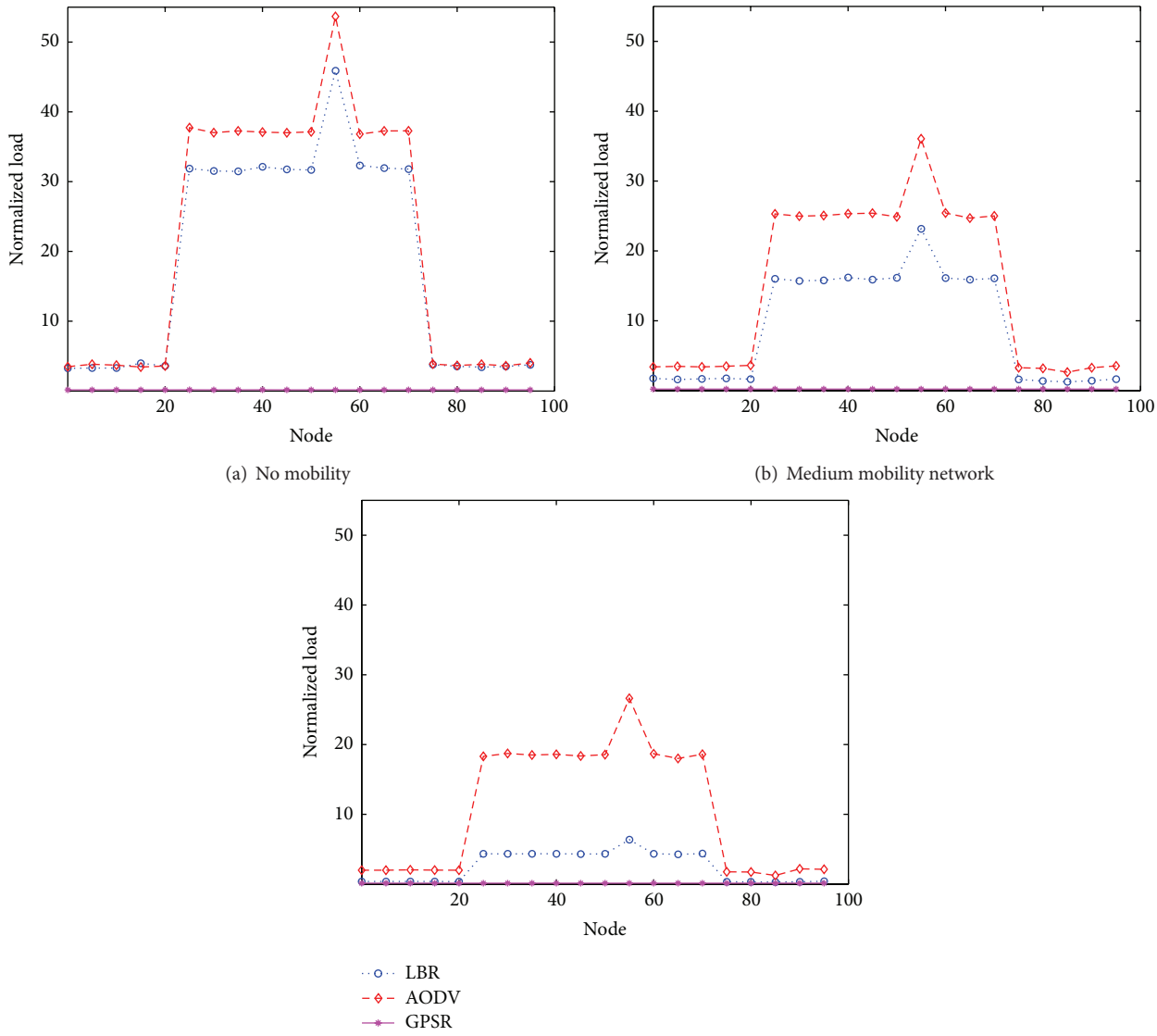


FIGURE 9: Average normalized routing load per node.

distribution contributes towards better lifetime and lower latency figures achieved by proposed scheme. Although per node load for GPSR is less compared to the other two protocols, it can be attributed to premature network partitioning and lesser transferred data. Therefore, results do not reflect a better load balance and the same is evident from very less network lifetime also.

6. Conclusion and Future Work

Sensor network routing protocols are traditionally designed for specific environment and applications to achieve efficiency and assumptions generally made by designers are rather strong limiting protocol application in generic scenarios. Secondly, mobility, in general, is used for data collection and resource provisioning only. With proliferation of embedded sensors in consumer devices, greater variety of applications and accompanied challenges would need to be addressed. In this changed scenario, application paradigm is likely to transform from application specific smaller scale to larger or global one. To address these challenges, generic protocols capable of handling wide ranging applications, device heterogeneity, and uncontrolled mobility would be needed. Proposed protocol utilizes mobility in novel manner and is deemed to address these new challenges. The scheme is energy efficient, load balanced, and congestion resilient and can handle variety in sensor mobility. The protocol is equally suitable for static, mobile sensor networks and can handle event driven as well as continuous traffic flows. Simulation results show that LBR outperforms minimum hop routing and greedy forwarding in terms of network lifetime, load balance, and data latency. The scheme has comparable results as far as throughput is concerned.

In this work, weight selection is performed using simulation method which may not be optimal to handle variety

in mobility and load across different applications. Therefore, weight selection based on fuzzy logic or analytical hierarchical process (AHP) may be studied as future work. Moreover, in case of mobile sensor networks, link quality may vary more rapidly compared to static networks, so addition of reliability to routing metric will improve protocol performance and may be another dimension for future research.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," *Ad Hoc Networks*, vol. 3, no. 3, pp. 325–349, 2005.

[2] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson, "People-centric urban sensing," in *Proceedings of the 2nd Annual International Workshop on Wireless Internet (WICON '06)*, ACM, August 2006.

[3] F. Bouabdallah, N. Bouabdallah, and R. Boutaba, "On balancing energy consumption in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2909–2924, 2009.

[4] G. Sara and D. Sridharan, "Routing in mobile wireless sensor network: a survey," *Telecommunication Systems*, 2013.

[5] M. di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: a survey," *ACM Transactions on Sensor Networks*, vol. 8, no. 1, article 7, 2011.

[6] E. Ekici, Y. Gu, and D. Bozdag, "Mobility-based communication in wireless sensor networks," *IEEE Communications Magazine*, vol. 44, no. 7, pp. 56–62, 2006.

[7] T. L. Guohong Cao, G. Kesidis, and B. Yao, "Purposeful mobility and navigation," in *Sensor Network Operations*, pp. 113–126, 2005.

[8] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002.

[9] W. Wang, V. Srinivasan, and K.-C. Chua, "Extending the lifetime of wireless sensor networks through mobile relays," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1108–1120, 2008.

[10] L. Shi, B. Zhang, Z. Yao, K. Huang, and J. Ma, "An efficient multi-stage data routing protocol for wireless sensor networks with mobile sinks," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1–5, December 2011.

[11] K. Tian, B. Zhang, K. Huang, and J. Ma, "Data gathering protocols for wireless sensor networks with mobile sinks," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–6, December 2010.

[12] W. Liang, J. Luo, and X. Xu, "Prolonging network lifetime via A controlled mobile sink in wireless sensor networks," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–6, December 2010.

[13] T. Wimalajeewa and S. K. Jayaweera, "A novel distributed mobility protocol for dynamic coverage in sensor networks," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, December 2010.

[14] Y. H. J. Jung and J. Cho, "A mobility-aware efficient routing scheme for mobile sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 974705, 7 pages, 2013.

[15] N. Nasser, A. Al-Yatama, and K. Saleh, "Mobility and routing in wireless sensor networks," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '11)*, pp. 000573–000578, May 2011.

[16] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 4, pp. 609–619, 2004.

[17] M. Younis, M. Youssef, and K. Arisha, "Energy-aware routing in cluster-based sensor networks," in *Proceedings of the 10th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (MASCOTS '02)*, pp. 129–136, 2002.

[18] D. Bradley and R. Uma, "Energy-efficient routing through weighted load balancing," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 31–37, 2012.

[19] C. Diallo, M. Marot, and M. Becker, "Link quality and local load balancing routing mechanisms in wireless sensor networks," in *Proceedings of the 6th Advanced International Conference on Telecommunications (AICT '10)*, pp. 306–315, May 2010.

[20] A. Nayyar, F. Bashir, and U.-U. Ubaid-Ur-Rehman, "Load based energy aware multimedia routing protocol (LEAR)," in *Proceedings of the 3rd International Conference on Computer Research and Development (ICCRD '11)*, vol. 2, pp. 427–430, March 2011.

[21] R. Kacimi, R. Dhaou, and A.-L. Beylot, "Load-balancing strategies for lifetime maximizing in wireless sensor networks," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, pp. 1–5, May 2010.

[22] B.-J. Kwak, N.-O. Song, and L. E. Miller, "A canonical measure of mobility for mobile ad hoc networks," in *Proceedings of the IEEE Military Communications Conference (MILCOM '03)*, vol. 2, pp. 1319–1324, IEEE Computer Society, Washington, DC, USA, October 2003.

[23] F. Ren, J. Zhang, T. He, C. Lin, and S. K. D. Ren, "EBRP: energy-balanced routing protocol for data gathering in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 12, pp. 2108–2125, 2011.

[24] G. Anastasi, M. Conti, M. di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: a survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009.

[25] T. R. Burchfield, S. Venkatesan, and D. Weiner, "Maximizing throughput in zigbee wireless networks through analysis, simulations and implementations," in *Proceedings of the International Workshop on Localized Algorithms and Protocols for Wireless Sensor Networks*, pp. 15–29, Citeseer, Santa Fe, NM, USA, 2007.

[26] B. Latré, P. de Mil, I. Moerman, N. van Dierdonck, B. Dhoedt, and P. Demeester, "Maximum throughput and minimum delay in IEEE 802.15. 4," in *Mobile Ad-Hoc and Sensor Networks*, pp. 866–876, Springer, New York, NY, USA, 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 579271, 16 pages
http://dx.doi.org/10.1155/2014/579271

Research Article

Wireless M-Bus Sensor Networks for Smart Water Grids: Analysis and Results

S. Spinsante, S. Squartini, L. Gabrielli, M. Pizzichini, E. Gambi, and F. Piazza

Department of Information Engineering, Marche Polytechnic University, 60131 Ancona, Italy

Correspondence should be addressed to S. Spinsante; s.spinsante@univpm.it

Received 4 November 2013; Revised 4 April 2014; Accepted 5 April 2014; Published 12 June 2014

Academic Editor: Joel J. P. C. Rodrigues

Wireless sensor network technologies are experiencing an impressive development, as they represent one of the building blocks upon which new paradigms, such as Internet of Things and Smart Cities, may be implemented. Among the different applications enabled by such technologies, automatic monitoring of the water grid, and smart metering of water consumptions, may have a great impact on the preservation of one of the most valued, and increasingly scarce, natural resources. Sensor nodes located along water pipes cannot rely on the availability of power grid facilities to get the necessary energy imposed by their working conditions. In this sense, an energy-efficient design of the network architecture, and the evaluation of Energy Harvesting techniques to sustain its nodes, becomes of paramount importance. This paper investigates the suitability of a Wireless Metering Bus-based solution for the implementation of smart water grids, by evaluating network and node related performance, through simulations, prototype design, and experimental tests, which confirm the feasibility and efficiency of the proposal.

1. Introduction

Wireless sensor networks (WSNs) are becoming a pervasive technology to monitor very different systems, from wide geographical areas to small-scale phenomena (as in body area networks). They are able to promptly detect particular events or working conditions and deliver related information to a specific destination. WSNs are currently deployed in many application fields, such as buildings construction, traffic monitoring, environmental analysis, healthcare assistance, weather forecast, and many others [1–3]. WSNs can improve the way people interact with surrounding environment, to increase their quality of life, and reduce the wasting of natural resources (water, oil, and gas), while optimizing their usage. A relevant task from this perspective is represented by *metering*, that is, the capability of measuring the amount of a certain resource, available at a specific grid point. In particular, companies in the utility sector providing electricity, gas, or water services are clearly interested in advanced metering infrastructures (AMIs), to improve their real-time monitoring over consumptions and optimize plants management.

While the application of smart metering and monitoring solutions is now well established in the field of electricity distribution and power management, the situation is much less mature in the field of water plants management and monitoring. On the other hand, the availability of almost real-time data about water consumptions could enable a number of actions, such as control of water consumptions in a district, at a given time; planning of maintenance operations; dynamic water flow rate management; dynamic in-pipes pressure management, to reduce failure rates and optimize water delivery at the users' outlets [4]. Obtained data can also help detecting leaks in water pipelines: indeed, a leak changes the hydraulics of the pipeline and therefore flow or pressure readings, after some time [5].

Water availability and consumptions represent critical issues that need specific tools to anticipate problems and resolve them proactively, to analyze data for better decisions, and to coordinate resources for operating effectively. In fact, water is needed and used, for more than drinking only. Global industry uses 20% of the world's water supply; in the US, it is 46%; in China, water usage by industry amounts to 25%; India is only about 5%. According to reports by the United Nations, agriculture is the largest consumer of freshwater, by far: about 70% of all freshwater withdrawals go to irrigated crops. Though it is a worldwide resource, water

is often treated as a regional issue. However, the way people use water at a regional level may have large-scale impact. With advances in technology, thanks to sophisticated sensor networks, smart meters, and deep computing and analytics, it is possible to manage water in a smarter way. Smart water grid infrastructures allow monitoring, measuring, and analyzing entire water ecosystems, from rivers and reservoirs to the pumps and pipes in citizens' homes, thus providing reliable, timely, and actionable view of water use.

Following the preliminary evaluations presented in [6–8], this paper analyzes the adoption of the Wireless Metering Bus (or in short WM-Bus) standard [9] in future smart water grids, for real time monitoring of water plants and leakages prevention. The M-Bus is a well-known field bus technology, specialized for transmitting metering data from gas, heat, oil, or other meters to a data collector. The WM-Bus represents the wireless implementation of the M-Bus standard; at the authors' best knowledge, the WM-Bus has been adopted sporadically in power grid and gas metering applications, to deploy AMI solutions, but it has not yet found a significant adoption in the design of *smart* (i.e., ICT-enabled) water distribution infrastructures. This is due partly to their still relatively scarce development, partly because of the adoption of different wireless techniques, often proprietary, or borrowed from other applications (such as ZigBee-like transmission techniques). Water distribution networks are typically set up as cyber physical systems (CPSs), with related sensing and actuating technologies for monitoring and maintenance operations, only till the hierarchical level of districts. Most of the water distribution systems totally miss dedicated ICT capabilities at the edge branches of the network, from the district-level node to the final meters, located at the customers' premises. A capillary network operating according to the WM-Bus standard can provide a viable and effective solution to implement an overlay ICT infrastructure that fills in the technology gap in the edge domain of water distribution networks. The goal of this paper is to outline an integrated approach, encompassing the transmission technology for the capillary network, the design of the smart devices to be applied on water meters, and the evaluation of the channel behavior affecting the link performances, to achieve a smart water grid reference design, which also takes into account the problem of powering the nodes and enabling reliable data transmissions. The main contributions of the present research work deal with a performance analysis of the different WM-Bus modes on a real embedded platform, according to a holistic approach addressing packet error rate, transmission coverage, and message collision probability, and with a comparative evaluation of the different WM-Bus modes from the point of view of battery lifetime and Energy Harvesting (EH) sustainability, in particular with reference to the smart water grid applicative context.

The paper is organized as follows: Section 2 discusses the state of the art in the field of smart water grid technologies and projects, whereas Section 3 focuses on the WM-Bus protocol selected for the proposed solution, its related message formats, and transmission profiles. In Section 4, a prototype development of the WM-Bus peripheral nodes and gateway for the smart water grid is presented. Section 5 deals with the results of preliminary experimental evaluations performed on the prototype nodes and discusses the possibility to involve Energy Harvesting modules, to supply the required energy. Finally, Section 6 concludes the paper.

2. Overview of Smart Water Grid Technologies

Most of the WSN implementations proposed in the technical literature rely on the ZigBee (or IEEE802.15.4 based) short range radio communication technique, over the 2.4 GHz band. In the context of smart metering, the first version of ZigBee Smart Energy (ZSE) profile was ratified and published, in 2008 [10]. It was driven by many market players and a variety of products were certified only a few months after the standard was finalized. ZigBee Smart Energy version 2.0 provides new capabilities, such as control of plug-in hybrid electric vehicles (PHEVs) charging, home area network (HAN) deployments in multidwelling units, such as apartment buildings, support for multiple energy service interfaces into a single premise, and support for any transport based on IETF IP compliant protocols, such as ZigBee IP specification. It is basically designed for energy-related scenarios and less general purpose than the W-MBus protocol. Other protocols for different kinds of short range wireless networks also exist, as the 6LoWPAN protocol [11], which supports IPv6 over IEEE802.15.4. These solutions foresee nodes equipped with IP stack: the design of the single node has to face an increased complexity (each node needs some processing capability, memory, and power supply), but the network management and data communication are made easier, by the adoption of the standard IP architecture and related protocols.

On the other hand, due to the lack of a common network layer, ad hoc solutions, not based on IP, require an application-layer gateway to communicate with other systems, or even the wider Internet. Gateways provide a mapping between different protocols, with a significant effort in functional and semantic translation. In this sense, complexity is shifted from the edge node to the core of the architecture. This approach may be valid when the application scenario requires minimizing complexity at the edge of the network infrastructure, where the available resources (processing, memory, and power) are very limited. Such a condition holds in the case of smart water grids: sensor nodes of minimal complexity are located at the boundaries of the water plants (i.e., at each user's premise, either urban or rural), where even power supply may be a concern, whereas concentrators (acting as gateways) may be installed in intermediate positions, where weaker constraints about power supply or processing resources hold. This is the reference architecture considered within this paper, in the framework of which the WM-Bus protocol is evaluated, as the selected solution for the transmission of metering and quality-related data, generated from sensor nodes located along water plants.

Research related to monitoring water distribution systems has increased in recent time. The availability of continuous monitoring is critical to detect any change in water quality or pipeline health, such as damages or water leakages.

Table 1: Operating modes of Wireless M-Bus.

*S*1	Unidirectional	In the stationary mode, the metering devices send their data several times a day. In this mode, the data collector may save power as the metering devices send a wake-up signal before transmitting their data.
*S*1-*m*	Unidirectional	Same as *S*1, but the data collector must not enter low-power mode.
*S*2	Bidirectional	Bidirectional version of *S*1.
*T*1	Unidirectional	In the frequent transmittion mode, the metering devices periodically send their data to collectors in range. The interval is configurable in terms of several seconds or minutes.
*T*2	Bidirectional	Bidirectional version of *T*1. The data collector may request dedicated data from the metering devices.

Several solutions based on WSNs for water grid are reviewed in [12], where the main operational challenges to face are evidenced. Among them, the need of suitable power supply for peripheral nodes is a critical factor. Most of the current solutions proposed in the technical literature, or available on the market, rely on batteries to feed the leaf nodes and sensors, which cannot be connected to the power grid. But the difficulties related to maintenance and substitution of batteries and their relatively short lifetime (with respect to the very long lifetime of the water grid pipelines) make such a solution not completely satisfactory or reliable. In [13], the authors outline the application of a WSN for monitoring a common water treatment process. The WSN is arranged in a simple star topology configuration, with the base station acting as the central hub; wireless communications are carried out by an 868 MHz RF transmitter. Issues related to power supply for the network nodes are not discussed. A WSN to monitor water distribution systems is presented in [14], and the main focus of the paper is on the need of defining a suitable model for the underground-to-above-ground communication channel, in order to properly estimate the performance of a wireless transmission from a sensor node to a gateway, and the related power requirements. The results provided by such a channel definition are then used to estimate the maximum operating range for any sensor node working according to the IEEE 802.15.4/ZigBee transmission standard, in the 2.4 GHz band.

Despite the challenges ahead, smart grid technology for water makes plenty of sense and deployments of new technology will be steady. Beyond improved metering, emerging solutions involve new algorithms to optimize water release in urban areas depending on the real needs [15, 16], new sensor capabilities for better leak detection [17, 18], enhanced monitoring of water quality, and the ability to better detect security threats to water systems. The drivers for smart grid technology in water are compelling: worldwide demand for water is expected to soar 40% from current levels, according to the 2030 Water Resources Group [19]; and losses from unmetered water total $14 billion in missed revenue opportunities each year, according to the World Bank [20]. These drivers will help fuel a move to smart technology solutions that promise more efficient water systems. Evidence of this trend continues to mount. A few examples include, among the others, Australia's Sydney Water that began deployment of high-efficiency meters to replace its aging stock; the three-year program will enable Sydney Water (which serves 4.6 million people) to eventually take advantage of automated and advanced metering technology. In England, Thames Water is extending a smart grid trial in the town of Reading to the city of London to better manage consumption and leakage. In Charlotte, NC, a public-private effort called Smart Water Now is taking place to measure consumption and improve efficiency; the city has partnered with private industries to collect information with the aim of lowering operational costs and improving sustainability.

In the next future, the adoption of WSNs in the field of water distribution monitoring and consumption metering will probably undergo a large market growth, if policy makers and national institutions will recommend or enforce the massive use of smart meters and monitoring systems, to control water distribution, detect leaks, and preserve water resources.

3. The WM-Bus Protocol

Short or medium range communication technologies must ensure minimum power consumption. ZigBee or other solutions based on IEEE 802.15.4 have been widely used for low-power sensor networks, but other protocols, such as the WM-Bus, have been recently proposed by the Open Metering Systems group [21] for metering scenarios. WM-Bus transceivers require low energy thanks to a low-overhead protocol, transmission-only modes (which do not require an idle receive phase), and long range sub-GHz transmission bands. While the first document (EN 13757-4:2005) prescribed the use of the 868 MHz ISM and 468 MHz bands, the later version (EN 13757-4:2011) added new transmission modes at 169 MHz, with lower data rates. The lower 169 MHz frequency band enables longer transmission range due to the inherently lower path losses, while the reduced data rates enable higher sensitivity for the receiver, allowing a reduction of the transmission power at the transmitter, or a longer transmission range, at a parity of the transmission power.

Based on the specific application, there are combinations of communication modes for data collectors and metering devices. These settings define the communication flow and the configuration of the radio channel. Table 1 lists the available communication modes.

The basic WM-Bus modes of interest are the following ones:

(i) *T mode*: frequent transmission mode (several times per second or per minute), 868 MHz, 100 kbps data rate from meter to gateway; in mode *T2* the transmitter requires an acknowledgment (ACK), different from *T1*;

(ii) *S mode*: stationary mode (several transmissions per day), 868 MHz, 32.7 kbps data rate; in mode *S2* the transmitter requires an ACK, different from *S1*.

Further, in the 169 MHz band, the standard also foresees the following modes:

(i) *Nc mode*: 169.431 MHz, 2.4 kbps data rate; *N2c* requires ACK, *N1c* does not;

(ii) *Na mode*: 169.40 MHz, 4.8 kbps data rate; *N2a* requires ACK, *N1a* does not;

(iii) *Ng mode*: 169.437 MHz, 38.4 kps data rate; it always requires ACK;

and submodes:

(i) *N1a-f*: one-way transmission; the node transmits on a regular basis to a stationary receiving point; single hop repeaters are allowed;

(ii) *N2a-f*: two-way transmission; the node transmits like *N1a-f*; its receiver is enabled for a short period after the end of each transmission and locks on if a proper preamble and synchronization word is detected.

The WM-Bus link layer is compliant with EN 13757-4:2011.10. It provides services that transfer data between PHY and application layer, generates outgoing CRC, and verifies CRCs for incoming messages. Further, the link layer provides WM-Bus addressing, acknowledges transfers for bidirectional communication modes, deals with WM-Bus frame formation, and verification of incoming frames. Two frame formats are foreseen, named *A* and *B*, identified by a specific preamble/synch sequence. The standard provides a number of predefined messages that are not used to carry application-specific data (that depend, for example, on the specific sensor used to monitor the grid), but to manage operational conditions. As an example, a three-way handshake is foreseen during the meter installation step, to enable the meter registration at the concentrator. Once registered, the meter automatically leaves the installation mode, whereas the concentrator requires manual intervention or timeout. Figure 1 shows the three-way handshake process featuring the meter installation step.

The advantages of the 169 MHz band with respect to the 868 MHz are implicitly related to the *narrowband transmission* concept. The greater the bandwidth, the greater the noise at the receiver input: so, with a signal bandwidth of 25 kHz or less, the *N* mode introduces much higher link budget and provides longer range solutions than the ones allowed at 868 MHz. A narrowband solution results in a radio performance improvement without significant problems, because the amount of data to be transmitted in a metering scenario is very low, thus avoiding bottlenecks that would slow down the entire network performance.

Table 2: Typical values of the path loss exponent *n*.

Environment	Path loss exponent, *n*
Free space	2
Urban area cellular radio	2.7 to 3.5
Shadowed urban cellular radio	3 to 5
In buildings line of sight	1.6 to 1.8
Obstructed in building	4 to 6
Obstructed in building	2 to 3

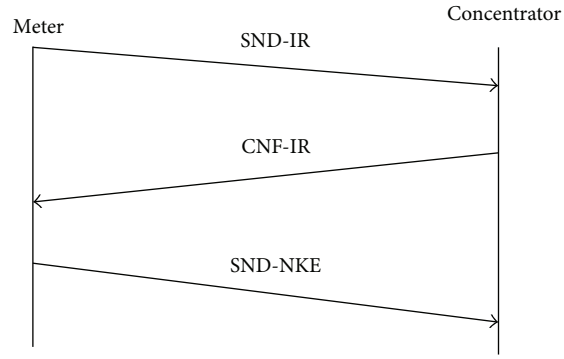


Figure 1: WM-Bus three-way handshake for the meter installation step.

The WM-Bus protocol also foresees the division of the available bandwidth in a number of channels. Up to six channels can be allocated for the data exchange between meter and concentrator, spaced by 12.5 kHz. If only one channel is not sufficient to meet the bandwidth requirements of the smart water grid, it is possible to consider the simultaneous use of more channels within the same interference domain. Such a frequency division multiplexing (FDM) capability is exposed to potential adjacent channel interference phenomena, but experimental tests have shown that the interference cancellation filters onboard of the WM-Bus transceivers are more than able to reduce the interfering signal power level 20 dB lower than the power level of the channel central frequency.

Another evident advantage of the 169 MHz band is related to the reduced path loss experienced by the propagating radio signal. The path loss exponent n in the generalized Friis' equation on propagation loss [22] varies according to the characteristics of the propagating environment, as detailed in Table 2. For $n = 2$ and $n = 3.5$ (free space and urban area propagation, resp.), the comparison between the *PL* values at 169 MHz and 868 MHz (at a parity of the antenna gains) confirms the better behavior of the radio transmission in the 169 MHz band, as shown in Figure 2. It is worthwhile to note that the two propagation conditions compared above are typical of a metering scenario, either applied to power or water grids, in which part of the grid is deployed in open areas (like rural ones), and part is located in urban environments. This higher communication distance does not require the use of repeaters, which are essential in the 868 MHz band

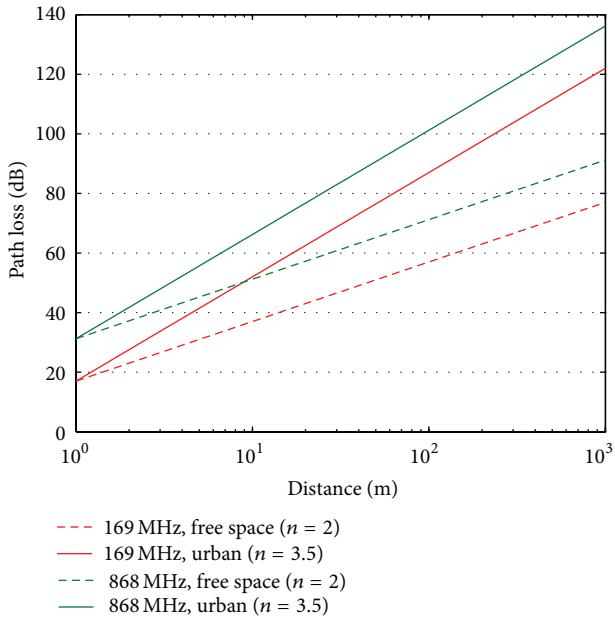


Figure 2: Path loss profiles at 169 and 868 MHz for different values of the index n.

to increase the communication range, so the overall cost of the network is lower. At the moment, the 169 MHz band is used only for remote control and smart metering, limiting the number of radio interferences. The main drawbacks related to the use of this frequency are the antenna size, too bulky for this kind of application (especially for the form-factor requirements of the peripheral nodes), and the lack of a mature technology. In this regard, a complete analysis of the network at 169 MHz is presented in the following sections.

The WM-Bus is able to achieve longer distance communication with respect to IEEE 802.15.4. Leaf nodes equipped with ZigBee/IEEE 802.15.4 transceivers can only cover a few tens of meters. Should greater distances be covered in ZigBee/IEEE 802.15.4 technology, a multihop data transfer strategy can be foreseen, according to which each leaf node can also act as a relay to another node, depending on the distance from the master node. Obviously, the ability of relaying depends both on the estimated amount of data each leaf node may generate and on the overall transmission capacity of the single ZigBee transceiver that must act as a relay: therefore it will be necessary to plan the placement of the nodes, once provided the amount of data to be transferred. Although the WM-Bus enables long ranges, it also provides relaying method to cover longer distances, by multihop techniques described in [23].

4. Prototype Implementation of a WM-Bus Solution for Smart Water Grids

4.1. Network Topology for the Smart Water Grid. The deployment of water grid networks is a basic step in the delivery of AMI services, besides enabling a continuous, prompt, and reliable monitoring of the water plants. In AMI scenarios, sensor nodes are usually displaced for domestic or industrial water accounting purposes, in a typical communication architecture that is organized according to a hierarchical topology, as the one shown in Figure 3. Peripheral WSN nodes are connected to master nodes, which act as gateways by collecting data and sending them to a centralized control and monitoring system, where data are stored and processed.

The hierarchical topology relies on the assumption that peripheral nodes are able to perform short or medium range radio transmissions, at low power consumption. On the contrary, master nodes are typically equipped with long range transmission capabilities, up to a geographic scale. Data generated from several peripheral nodes are collected and organized by each master node, so that they can be delivered to the central monitoring unit, where they are processed by suitable algorithms, to identify and locate possible faults along the pipes, in a real time fashion. Each network element, especially the numerous remote nodes, shall be designed in such a way as to keep global maintenance operations at a minimum.

The architecture proposed for the monitoring network assumes master nodes equipped with GSM/GPRS modules. In fact, master nodes that collect monitoring-related data from multiple leaf nodes must be able to deliver them to the central unit, typically along distances of a few kilometers. Long range transmissions generated by master nodes are sustained by power supplied from the grid, or from solar cells; short/medium range communications, as those of leaf sensor nodes, must require the minimum power consumption, usually provided by batteries. At the same time, in order to limit the overall costs related to the water grid deployment, the number of peripheral nodes should be somehow kept small. Based on the analysis of these requirements, the WM-Bus standards is assumed as a good tradeoff among the power needs due to radio transmission and the minimum coverage range necessary to limit the number of sensor nodes that shall be located along the water grid.

4.2. Sensor Node Implementation. Through the use of a suitable Texas Instruments (TI) development kit [24], a prototype node is implemented, which is useful to evaluate the performance of the WM-Bus protocol, in a possible and realistic smart water grid scenario. The kit provides a complete development platform for TI's Sub-1 GHz devices, comprising two SmartRF Transceiver Evaluation (TrxEB) boards, which include an MSP430F5438 MCU [25], a USB interface, a dot matrix LCD, and other useful features. A fully integrated single-chip radio transceiver, the CC1120 [26], is used, mainly intended for the ISM and SRD frequency bands, in the ranges 164–192 MHz, 410–480 MHz, and 820–960 MHz. The RF implementation guarantees good performance in terms of covered area, and power consumption. The output power can be increased up to +27 dBm (500 mW) by equipping the module with a power amplifier. The available application note [27] has been fruitfully used to develop WM-Bus applications. In the experimental tests performed, one of the boards shown in Figure 4 acts as sensor node and the other as gateway.

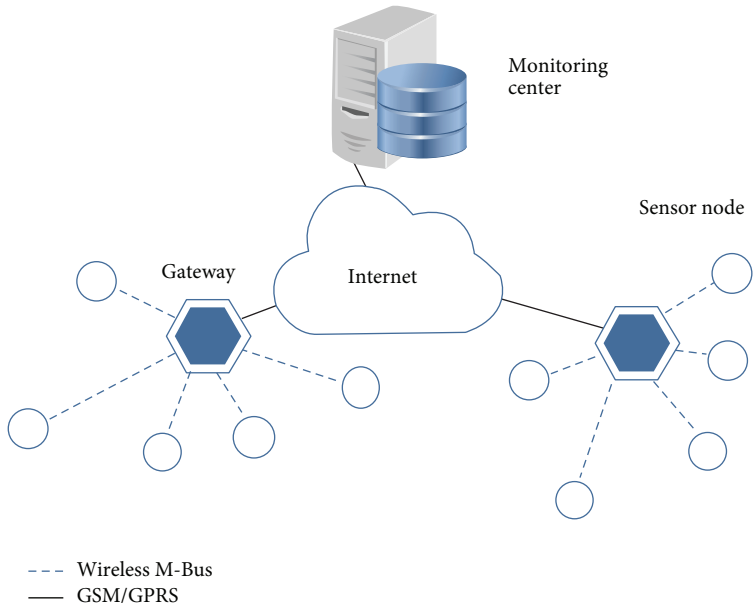


Figure 3: Sample WSN architecture for the water grid.

Figure 4: TI boards used to implement prototype WM-Bus sensor node and gateway.

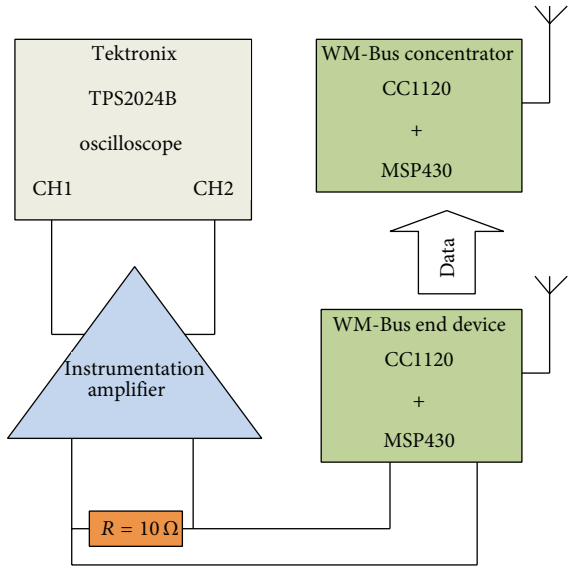


Figure 5: Measurement setup to evaluate WM-Bus related currents.

4.2.1. Estimation of WM-Bus Power Requirements. The power requirements of the WM-Bus protocol have been estimated by measuring the related currents. It is not a trivial task, when currents are in the sub-mA range, as in low power modes (LPMs): the device awakens from sleep only a few times a day for transmission; hence the main current draw is due to sleep currents. To measure the current at steady state, a current probe (*picoamperometer*) could be added in series to the supply circuit. However, to measure peak and transient current draws, an oscilloscope must be used. This requires reading a voltage across a precision resistor, by applying Ohm's law. For sub-mA currents, the voltage must be amplified by a precision amplifier [28] and realized as shown in Figure 5. The amplifier used in experiments is an AD621 from analog devices, which provides a 100 V/V gain up to 200 kHz, whereas the load is a 10 Ω resistor, yielding a gain factor equal to 1000.

4.2.2. MSP430 Power Consumption. In the active state, and in the sleep mode LPM3 (low power mode 3) of the MCU, the steady state current draw has been measured by using a picoamperometer. To achieve the minimal power consumption, the MCU and its peripherals must be initialized properly and/or shut down. The measured current draw in LPM3 amounts to 2.7 μA, very close to the figure of 2.1 μA reported in the device datasheet. Other possible LPMs were

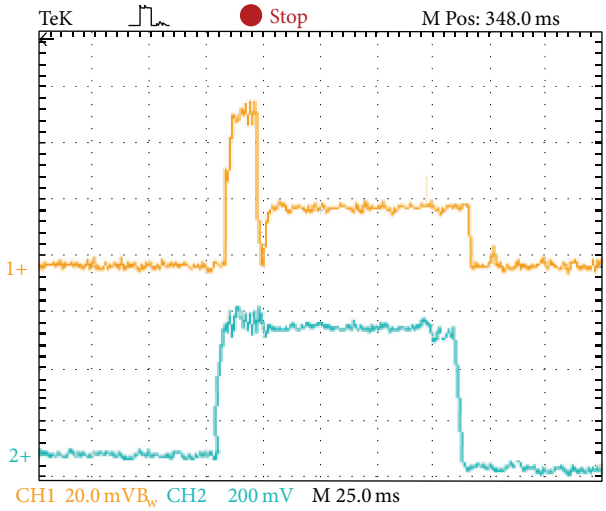


Figure 6: Current profile for the CC1120 transceiver (top) and the MSP430 MCU (bottom), during the transmission of a WM-Bus message in the N2g 38.4 kbps mode. Please note that the oscilloscopes read the transceiver current values from a 1 Ω resistor, while the MCU values are amplified and transformed into voltage signals, by the setup of Figure 5.

discarded as not adequate for the application scenario of interest. For instance, LPM0 keeps all the system clocks running, thus yielding a current draw of about 85 μA; LPM4 disables the CPU and all the clocks, yielding the lowest possible current consumption, around 0.1 μA. However, the latter mode requires an external circuitry to transmit a wake-up signal, which is out of the current research focus.

Differently from the LPMs, the measure of active MCU current draw gives many degrees of freedom, for example, choice of the clock frequency, FLASH or RAM use, selection of active peripherals, CPU instructions, and so forth; hence the measured value must be considered indicative of average MCU energy requirements. In the water metering application scenario no particular computational power or processing speed is required. As a consequence, the clock frequency was kept to the lowest value of 1 MHz; all the peripherals were disabled, together with the clock signals (as a consequence, user-specific application will require accurate profiling of peripherals power requirements). The CPU cycled in a while loop, performing memory reading and writing and arithmetic instructions, to simulate a moderate CPU workload. In this case the measured MCU consumption is 370 μA, very close to theoretical datasheet figure of 330 μA.

4.2.3. CC1120 Power Consumption. To measure transient and peak current, a Tektronix TPS2024B oscilloscope was connected to the setup shown in Figure 5. The transceiver supply current ran through a 5 Ω shunt resistor. Values read are thus amplified and transformed from amperes to volts. Typical time profiles of current drawn by the transceiver (top), and the MCU (bottom), are shown in Figure 6, whereas Table 3 lists the measured current consumptions. The protocol requires a transmission phase, corresponding to the initial high current peak featured by the transceiver, followed by a longer receive phase, providing the subsequent flat in the transceiver current measure. The MCU requires an almost constant current during the active phase, with a slightly higher average value during the transceiver transmission phase. Before and after its activation, the MCU is in a sleep state, and the transceiver is idle.

Table 3: Measured current consumptions for the CC1120 transceiver.

Transmission power [dBm]	TX current consumption [mA] at 169 MHz	TX current consumption [mA] at 868 MHz
+15	55	50
+14	51	46
+10	43	35
+5	33	30
+0	28	27
−5	26	25
−10	25	24
−16	24	23

4.3. Protocol Energy Requirements. The energy requirements exhibited by the protocol are evaluated for a single communication cycle, on a bidirectional link, at an operational frequency of 169 MHz and 868 MHz respectively, assuming the hardware and the firmware stack described above. All the transmissions are performed at a 15 dBm radiating power. To perform this test, Wireless M-Bus packets of 96 bytes are considered. This value is required to calculate the transmission period at the different data rates, according to

$$T_{\mathrm{TX}} = \frac{8 \cdot \mathrm{Packet\_Length\,[bits]}}{\mathrm{Data\_Rate\,[kbps]}}. \quad (1)$$

4.3.1. Energy Requirements at 169 MHz. The WM-Bus protocol prescribes several data rates for the *N* mode. One of these, *N2g* 38.4 kbps, is shown in Figure 6. In the example shown, the TX phase lasts 20 ms, with a current consumption by the transceiver of 55 mA. The RX cycle lasts 90 ms, with a current consumption of 22 mA. The MCU is active during the total communication time and exhibits a current draw of approximately 490 μA. This is 120 μA higher than the value provided by the picoamperometer, as reported in Section 4.2.2, because the SPI bus is now enabled to communicate with the CC1120 device. The LPM3 current also increases to 3 μA because of the SPI activation. In the example under examination, the overall energy requirement for one communication cycle amounts to

$$\begin{aligned} E_{N2g} &= 490\,\mu\mathrm{A} \cdot 20\,\mathrm{ms} \cdot 3\,\mathrm{V} + 55\,\mathrm{mA} \cdot 20\,\mathrm{ms} \cdot 3\,\mathrm{V} \\ &\quad + 490\,\mu\mathrm{A} \cdot 90\,\mathrm{ms} \cdot 3\,\mathrm{V} + 22\,\mathrm{mA} \cdot 90\,\mathrm{ms} \cdot 3\,\mathrm{V} \quad (2) \\ &= 9.3\,\mathrm{mJ}. \end{aligned}$$

Similar evaluations can be obtained for the *N2a* and *N2c* modes, operating, respectively, at 4.8 kbps and 2.4 kbps

Table 4: Device current draws.

	Operating mode	Measurement	Data sheet
MCU (MSP430)	Active	370 μA	330 μA
	Sleep	2.7 μA	2.1 μA
RF transceiver (CC1120)	Tx (868 MHz at 15 dBm)	50 mA	50 mA
	Tx (169 MHz at 15 dBm)	55 mA	54 mA
	Tx (169 MHz at 27 dBm)	140 mA	140 mA
	Rx	22 mA	22 mA
	Sleep	0.3 μA	0.5 μA

data rates. The low data rates increase the communication time needed, at a parity of the amount of data to transfer. The energy requirements increase too and reach 27.5 mJ and 49.8 mJ, respectively. Low data rates, however, also increase the receiver sensitivity, thus enabling longer communication ranges.

4.3.2. Energy Requirements at 868 MHz. As for the 169 MHz modes, the energy requirements are also discussed for the 868 MHz modes. The highest data rate mode is the 100 kbps *T* mode. The high data rate allows for very short transmission and receive time (10 ms and 78 ms, resp.), so that the energy requirements of the *T* mode are the lowest over all the WM-Bus modes, 6.7 mJ. Furthermore, the transceiver at hand has slightly lower power consumption for the same radiating power (15 dBm) at 868 MHz. It must be noticed, however, that the communication range will always be shorter than in the 169 MHz modes, due to higher path loss.

The *S* mode sends data at a rate of only 32.768 kbps. The energy consumption for a single transmission cycle is 9.7 mJ, slightly higher than the most energy efficient mode at the 169 MHz carrier frequency.

5. Simulations and Experimental Evaluations

5.1. Battery Lifetime. In order to estimate the power constraints of a leaf node in the smart water grid, a battery lifetime model is considered. It is assumed to power the board with a typical 3 V Lithium CR2354 battery (Panasonic lithium coin data sheet: http://www.alliedelec.com/search/product-detail.aspx?sku=70197003.), featuring a charge capacity equal to $C_b = 560$ mAh. The device current draws have been obtained by aforementioned measurements and are reported in Table 4.

Classical equations used to describe battery lifetime [29] can be applied also in the WSN case study, by adding a duty cycle factor and including the possibility to have different current values in compliance with the task executed by the sensor node processor. In particular, each current draw can be weighted by the activity time related to the corresponding transmission, receiving and sleep tasks, as described above. The following values hold, with one transmission per hour in the *S* operating mode:

(i) transmission: $T_{\mathrm{TX}} = 23$ ms, therefore we have $\Delta_{\mathrm{TX}} = 0.023/3600 = 0.000639\%$;

(ii) receiving: $T_{\mathrm{RX}} = 100$ ms, therefore we have $\Delta_{\mathrm{RX}} = 0.1/3600 = 0.00278\%$;

(iii) sleep: this phase is active for the majority of time; therefore it is reasonable to assume that $\Delta_{\mathrm{SLEEP}} \approx 1$;

(iv) MCU consumption: 370 μA;

Table 5: Battery lifetime for different WM-Bus operating modes.

Operating Mode	Battery Lifetime
S2	≈15 y 8 m
T2	≈15 y 10 m
N2a 2.4 kbps	≈13 y 1 m
N2a 4.8 kbps	≈14 y 5 m
N2g 32.8 kbps	≈15 y 7 m

where $i = \{\mathrm{TX}, \mathrm{RX}, \mathrm{SLEEP}\}$ denotes the possible state and Δ_i the duty cycle factor. The *battery lifetime* is then calculated for the *S* and *T* operating modes at 868 MHz, as well as for the *N* mode (for the two different data rates considered) at 169 MHz. It is assumed that a transmission is executed every 6 hours, which represents a reasonable time for metering scenarios. All reported values in Table 5 are relative to 15 dBm of irradiated power.

Figure 7 reports the battery lifetime curves in dependence on the time period between two consecutive transmissions (varying this value from a minimum of 1 hour, i.e., 24 per day, to a maximum of 6, i.e., 4 per day). It is evident that when the daily transmissions are less frequent the battery lifetime is very low and varying with the number of transmissions, therefore mostly dependent on the sleep consumption. In contrast, when the transmission frequency is higher, the power required to sustain the TX and RX phases is much more relevant. Moreover, the WM-Bus at 169 MHz is more demanding with respect to the 868 MHz counterpart, as expected from the energy measurements discussed above.

Moreover, to evaluate the energy requirement of a more performing radio performance setup, the CC1120 transceiver has been equipped with a power amplifier to reach an irradiated power of 27 dBm (500 mW), with a consequent current consumption increase up to 140 mA. To analyse the battery lifetime in this case, a comparison between 15 dBm and 27 dBm has been carried out in the N_a mode case study, as shown in Figure 8.

Assuming also in this case a transmission every 6 hours, the resulting battery lifetime for the 27 dBm transmission is

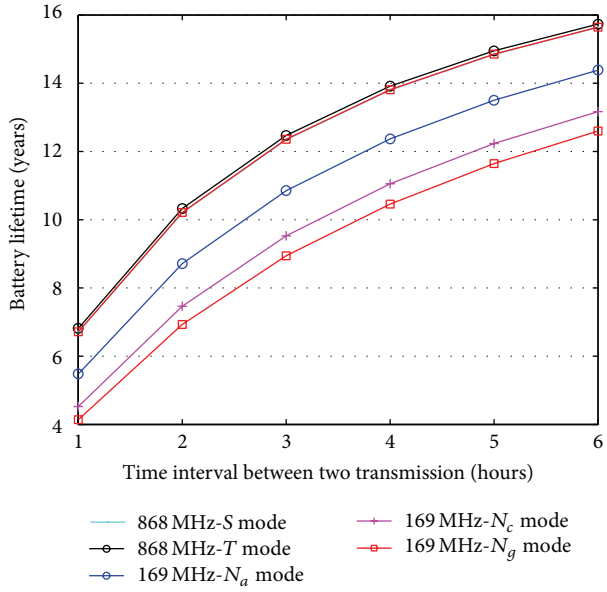


Figure 7: Battery lifetime versus the time period between two consecutive transmissions, for different frequencies and operating modes.

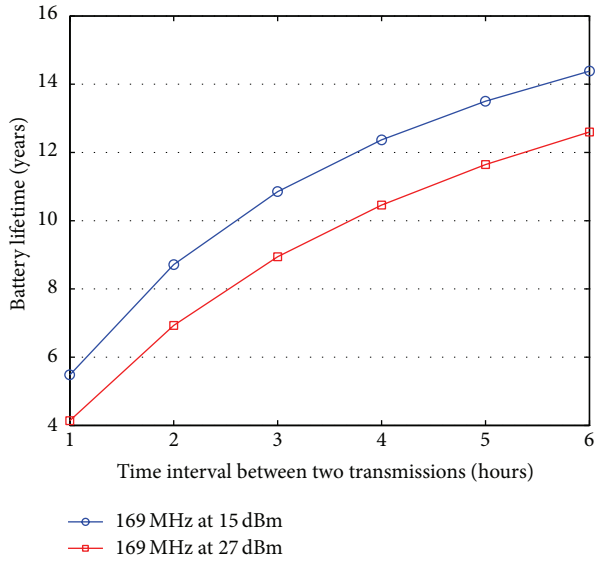


Figure 8: Battery lifetime versus the time period between two consecutive transmissions at 169 MHz, at 15 and 27 dBm.

about 12 years and 7 months, almost 2 years less than the 15 dBm case.

5.2. Sensor Node Powering by Energy Harvesting. The possibility to involve Energy Harvesting (EH) modules to supply the required energy for smart water metering sensor nodes is discussed in this section. Such an issue has a strong relevance in the addressed application scenario, since the deployed WSN could be characterized by nodes which are not easily accessible and therefore the eventual substitution of batteries over time can have an impact in costs management.

Among the various EH boards available on the market, the EH300A module (EH300A, Advanced Linear Devices Inc. available at http://aldinc.com/pdf/EH300ds.pdf) has been taken as reference and used in our evaluations. The EH300A module supplies and stores power for inputs higher than 4 V and 500 nA. Its storage capacity is 100 μAh. Inputs with values below this threshold do not activate the EH300A circuitry; thus, a solar panel providing maximum 200 mA at 5 V has been chosen to feed the EH300A module. The EH300A outputs a signal at the same voltage as the input signal, in this case 5 V. The TrxEB, however, can only accept input power in the range 3–3.5 V; hence a voltage regulating circuit has been cascaded between the EH300A and the TrxEB board, based on the LM317 integrated component. Under good solar radiation the TrxEB is able to work continuously without issues. However, with no solar power and fully charged storage buffers, the TrxEB was able to only sustain a second of continuous transmission as if the load required 600 μA. This is mainly due to additional circuitry on the TrxEB which cannot be disabled. That is why the related energy effort has not been accounted for when measuring the MCU and transceiver current draws.

5.2.1. Affordable Transmission Cycles. In order to evaluate the number of allowed transmission cycles in optimized HW, the aforementioned consumption values related to current draws only due to MCU and transceiver active and sleep states are considered. Thus, knowing the battery capacity C_b and the current C_i required by each state in a transmission cycle, the number of available transmission cycles is given by

$$N = \frac{C_b}{\sum_{i=1}^{N} C_i}. \quad (3)$$

Assuming to have a transmission cycle every 6 hours, as in Section 5.1, and a power irradiation of 15 dBm, the following are the values for most relevant S mode variables:

(i) energy storage capacity: $C_b = 100\ \mu$Ah;

(ii) cpacitors discharge cut-off threshold: $t_c = 50\%$;

(iii) total current consumption for TX: $I_T = 50.370$ mA;

(iv) average TX time: $T_T = 30$ ms;

(v) total current consumption for RX: $I_R = 22.370$ mA;

(vi) average RX time: $T_R = 78$ ms;

(vii) sleep time: $T_s = 21600$ s = 6 h.

The discharge cut-off threshold of capacitors limits the storage capacity: for instance, with $t_c = 0.5$, the available storage capacity is 50 μAh. In the light of this, the number of available transmission cycles in the S operating mode can be calculated as follows:

$$N_S = \frac{50\ \mu\text{Ah}}{\sum_{i=1}^{N} C_i} \approx 2. \quad (4)$$

Approximately the same result stands for the other WM-Bus modes. Figure 9 reports the N values calculated for all the

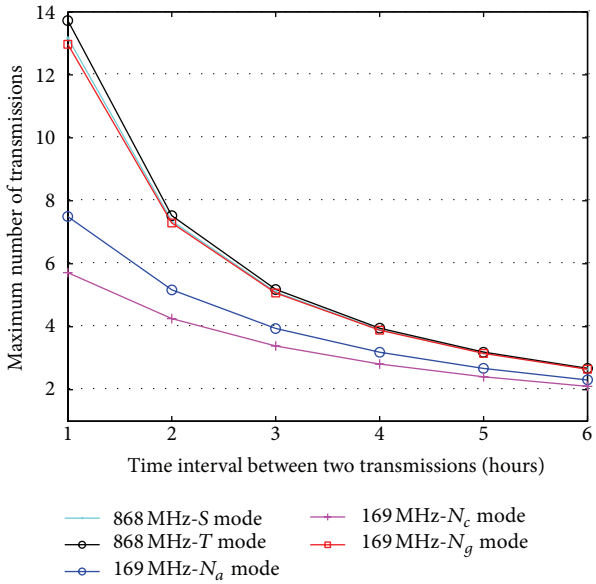


Figure 9: Maximum number of transmissions versus the time period between two consecutive transmissions. All possible operating modes at 15 dBm are evaluated.

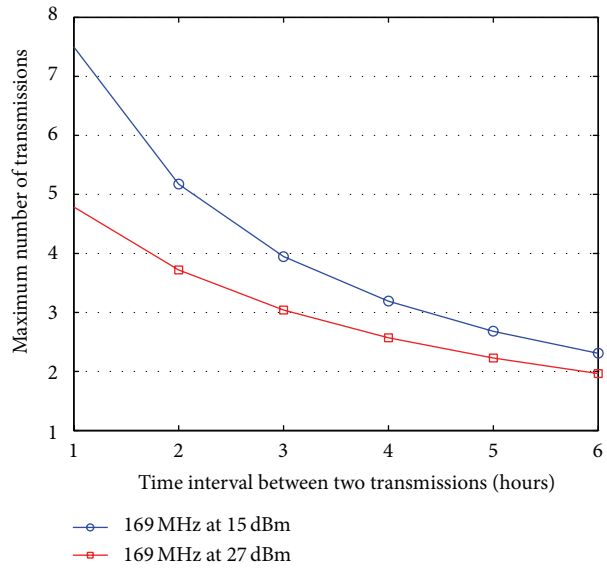


Figure 10: Maximum number of transmissions versus the time period between two consecutive transmissions. The N_a operating mode at 169 MHz, at 15 dBm and 27 dBm, is evaluated.

WM-Bus modes and parameterized with respect to the sleep time T_s between two consecutive transmissions.

Similarly to what is done in previous section, a comparison between 15 dBm and 27 dBm has been carried out from this perspective too, again in the sole N_a mode case study. Results are shown in Figure 10.

This evaluation proves the theoretical possibility of supplying power with intermittent energy harvesting to a sensor node implementing the WM-Bus protocol (both at 15 dBm and 27 dBm), with time between transmissions of 1 to 6 hours, which are typical of the metering scenario.

5.2.2. Considerations on Water Kinetic Energy as Energy Harvesting Source. Different energy harvesting sources can be effectively employed in the addressed application scenario. Solar power can be effective but the inherent energy intermittency and logistic problems in placing solar panels for meters in buildings suggest finding some more suitable solutions. That is why water kinetic energy can be used on purpose and a negligible amount of this kinetic energy can be converted into electric energy to power the sensor nodes: in this section a brief analysis of its suitability is discussed.

Water kinetic energy can be converted to electrical energy by a small dynamo turbine placed in contact with the flowing water (e.g., in a duct). Even in small ducts or at small flow rates (in the order of approximatively 10 L/min), this device allows for the harvesting of a moderate quantity of energy and proves very cheap in production and low in embodied energy with respect to solar panels, for instance. Preliminary tests were conducted with a commercial turbine from Seeed Studio Works (http://www.seeedstudio.com/depot/36v-micro-hydro-generator-p-634.html?cPath=155) for microhydropower generation. The turbine is 80 × 81.4 × 43.8 mm in size, with a nominal output of 3.6 V and 300 mA for water flow ranging from 1.5 to 20 L/min. It must be noted that a small capacity 300 mAh buffer battery is provided inside the turbine to achieve a continuous supply of energy and regulated voltage output.

The maximum nominal output power of the turbine is 1 W, enough for powering a sensor node with the power requirements discussed in the previous section when continuous water flows at sufficient speed in the duct. Even with intermittent water flowing (of the entity of those in a domestic, industrial, or public building), the EH still satisfies the SN needs. To recharge a typical water closet, for instance, 7 L of water is required to flow in 30 s on average. This is enough to supply approximately 1 W for 30 s, equivalent to 30 J of energy that can be stored in the battery and ideally cover, for example, 120 days of transmission with WM-Bus Nc mode (the most expensive energy, as highlighted above) at a rate of 1 transmission every 10 minutes [6].

5.3. Experimental Tests and Results on WM-Bus Transmissions. As a first estimate of the WM-Bus transmission performances, the amount of error packets for different values of the received power has been evaluated, through the packet error rate (*PER*) figure.

In order to test the transmission performances in presence of a controlled attenuation, the WM-Bus transceivers have been connected by cables (thus replacing their antennas), through two different lab attenuators: 355C attenuator by HP, which provides an attenuation ranging from 0 to 12 dB by steps of 1 dB; 355D attenuator by HP, which provides an attenuation ranging from 0 to 120 dB, by steps of 10 dB. Both the attenuators are able to operate in the VHF band, from DC to 1 GHz frequency. In order to check the level of received power, a suitable registry of the CC112X transceiver has been

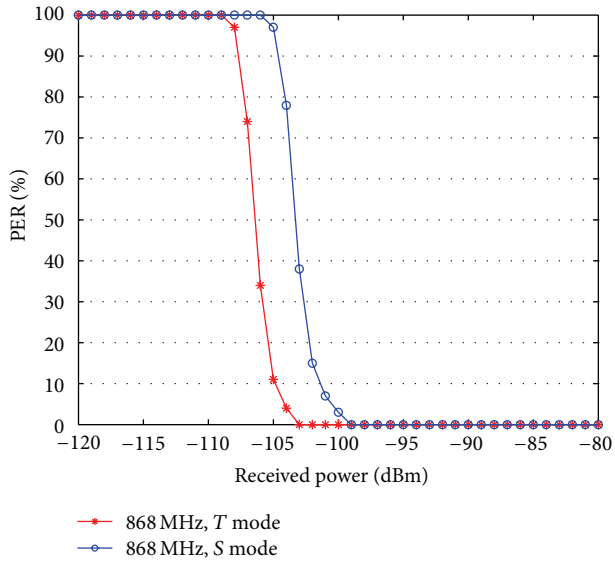


Figure 11: PER evaluation in the 868 MHz band.

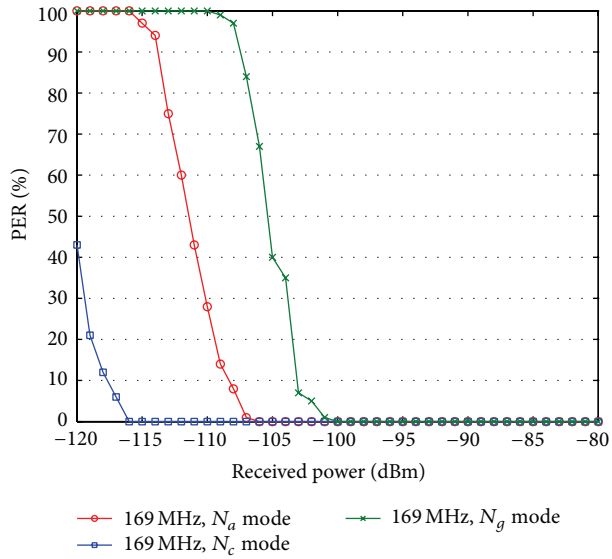


Figure 12: PER evaluation in the 169 MHz band.

monitored, the *RSSI* (received signal strength indication) one, which is used to store the signal power value each time new data are received. In the experiments, the *PER* has been evaluated around the sensitivity threshold of the transceiver, by varying the attenuation stepwise by 1 dB and for a total amount of 1000 data packets transmitted at each test.

Two transmission modes, T and S, have been tested for the 868 MHz operational band, as shown in Figure 11; N_g, N_a, and N_c modes have been tested for the 169 MHz band, as reported in Figure 12.

In order to explain the results obtained by the experimental tests, it is necessary to point out that when noise increases, also the receiver sensitivity increases, that is, the minimum power the receiver can detect. Noise may be due to several sources and phenomena; however, among the most significant contributions, the equivalent noise power at the device input N, which depends on bandwidth B, shall be accounted for. The relationship among the input receiver sensitivity S_{in} and the bandwidth B may be expressed as follows:

$$S_{\text{in}} = \text{NF} + N(B) + \frac{E_b}{N_0}, \tag{5}$$

where NF is the noise figure of the receiver and E_b/N_0 is the minimum signal-to-noise ratio needed to process a signal. Looking at Figure 11, how the error probability, which is related to the receiver sensitivity, also depends on the transmission data rate is shown. As a matter of fact, at a 100 kbps data rate (T mode), the sensitivity is around −106 dBm, whereas at a lower data rate (32.768 kbps, S mode), sensitivity decreases to around −109 dBm. In a similar fashion, sensitivity decreases with decreasing data rates also in the 169 MHz operational band, as shown in Figure 12, where sensitivity is around −110 dBm, −116 dBm, and −120 dBm for data rates of 38.4, 4.8, and 2.4 kbps, that is, respectively, N_g, N_a, and N_c modes.

Sensitivity depends on bandwidth more than on frequency; this relationship is further evidenced by the comparison between S mode at 868 MHz and N_g mode at 169 MHz, the data rates of which are, respectively, 32.768 and 38.4 kbps. As expected, the two modes show similar behaviors, but the S mode provides slightly worse sensitivity than the N_g mode.

5.4. Evaluation of Maximum Transmission Coverage. The evaluation of WM-Bus transmission coverage has been performed either by simulation, resorting to classical analytical models used to estimate the signal attenuation in different environments, or by practical measurements on the field, by using the same prototype devices discussed in the previous subsection.

In order to get a simulated estimation of the presumable maximum transmission range, the Okumura-Hata model [30] has been applied, due to the frequency range in which it can be used, [150; 1500] MHz, which includes the WM-Bus operational frequency bands. Three different scenarios are available in the model; in each case, the corresponding equation for loss estimation accounts for the height at which both the transmitting and receiving antenna are located and for the specific position of the receiver (such as indoor or outdoor), through suitable constant coefficients. For the purposes of the present research work, the open area model has been considered at first, thus simulating sensor nodes location corresponding to water pipes in rural areas. The signal loss in open areas, L_O, is given by

$$L_O = L_U - 4.78\log^2(f) + 18.33\log(f) - 40.94, \tag{6}$$

where f is the operating frequency (in MHz) and L_U denotes the signal loss expression for urban areas (in dB), given by

$$\begin{aligned} L_U = {} & 69.55 + 26.16\log(f) - 13.82\log(h_B) \\ & - C_H + [44.9 - 6.55\log(h_B)]\log(d), \end{aligned} \tag{7}$$

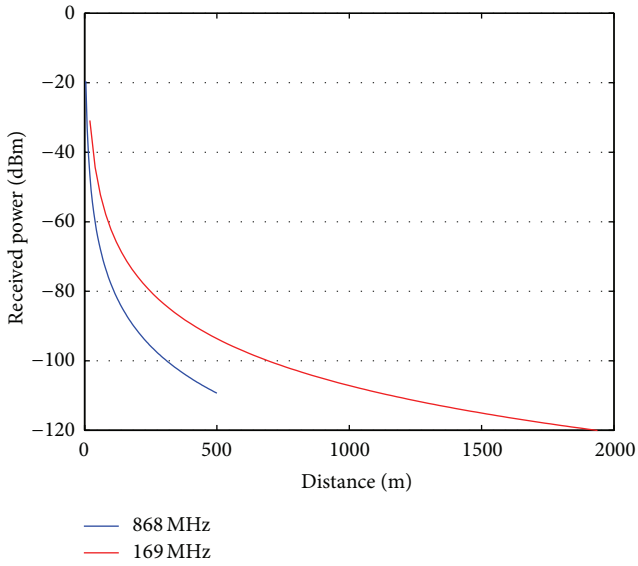


Figure 13: Received power according to the Okumura-Hata open area model, at 868 and 169 MHz.

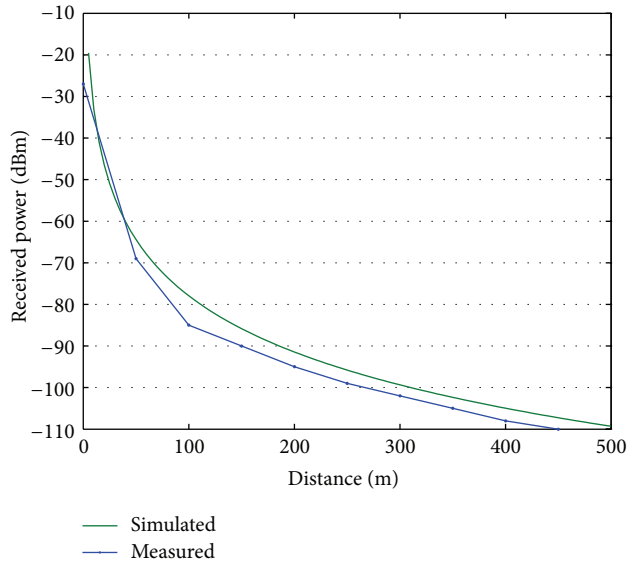


Figure 14: Received power according to the Okumura-Hata open area model and field measurements, at 868 MHz.

with h_B being the fixed station antenna height, d the transmitter-receiver distance in Km, and C_H a corrective factor for the antenna, given by $C_H = 0.8+(1.1\log(f)-0.7)h_M-\log(f)$, where h_M denotes the height of the mobile terminal antenna.

By assuming that L_O is equal to the difference between the transmitted power and the receiver sensitivity (so that the maximum acceptable attenuation is considered) and $h_B = h_M = 1$ m as a reasonable assumption, it is possible to compute the distance d covered by the WM-Bus communication, a 868 and 169 MHz. The results obtained are shown in Figure 13, where sensitivity is set to −109 dBm at 868 MHz and −120 dBm at 169 MHz.

The simulation shows a maximum coverage distance of 500 m at 868 MHz and up to almost 2 Km at 169 MHz. The WM-Bus 169 MHz operational band seems more suitable to ensure adequate coverage along the water pipes infrastructure, by means of a limited number of sensor nodes displaced, thanks to the greater transmission range supported. Simulated performances have been confirmed for the 868 MHz band, by experimental tests executed on the field, locating the WM-Bus transmitter and receiver in an open environment (an empty stadium park area). Figure 14 shows a good agreement among the expected and measured received power values, at different distances.

On the contrary, a strong disagreement among simulated and measured power values was evidenced at 169 MHz. This misbehavior was actually due to the helix antennas first adopted at 169 MHz in the prototype, which were far from ideal ones. The problem was solved by resorting to better designed and refined antennas, able to provide the expected performance.

5.5. WM-Bus Signal Coverage and Node Location. In order to compare the expected performance of the WM-Bus protocol at 868 and 169 MHz, a reference urban scenario has been simulated, in which we assume to have 4 buildings (each 30×30 m^2 large), in an area of 100×100 m^2. The Okumura-Hata model used to estimate signal propagation does not apply to distances shorter than 1 Km; as a consequence, for the simulated scenario, we adopt Friis' equation with $n = 2$, and introduce a fade margin in the range $20 \div 30$ dB, to account for signal attenuation due to obstacles, and possible multipath effects. In detail, in order to consider a worst case scenario, simulations are performed by assuming a 30 dB attenuation margin. For each building, 10 active sensor nodes are assumed, located at the bottom of the building (as usually happens for water meter in urban installations); two scenarios are considered, featuring one and two concentrators, respectively. For each scenario, the concentrator may be located at the center of the 100×100 m^2 area, or at a vertex position.

Figures 15 and 16 show the distribution of the received signal power, when a single concentrator is used, at center and vertex position, respectively, for 169 and 868 MHz operational frequencies. Due to the geometry of the scenario and the signal propagation effects, the better performance is obtained when a single concentrator is located in central positions and works at a frequency of 169 MHz.

Similar results are obtained when simulations account for two concentrators, located at the center of the building area, or at two vertices, as shown in Figures 17 and 18. Better behavior of the WM-Bus transmission is obtained at 169 MHz; in this case, however, the position of the concentrators does not affect signal coverage significantly, as the sensor nodes are almost uniformly covered.

5.6. Evaluation of Message Collision Probability. In order to evaluate the risk of collisions among asynchronous data transmissions generated by different sensor nodes, towards the same concentrator, it is assumed that the WSN upon which the smart water grid is built comprises n nodes, which

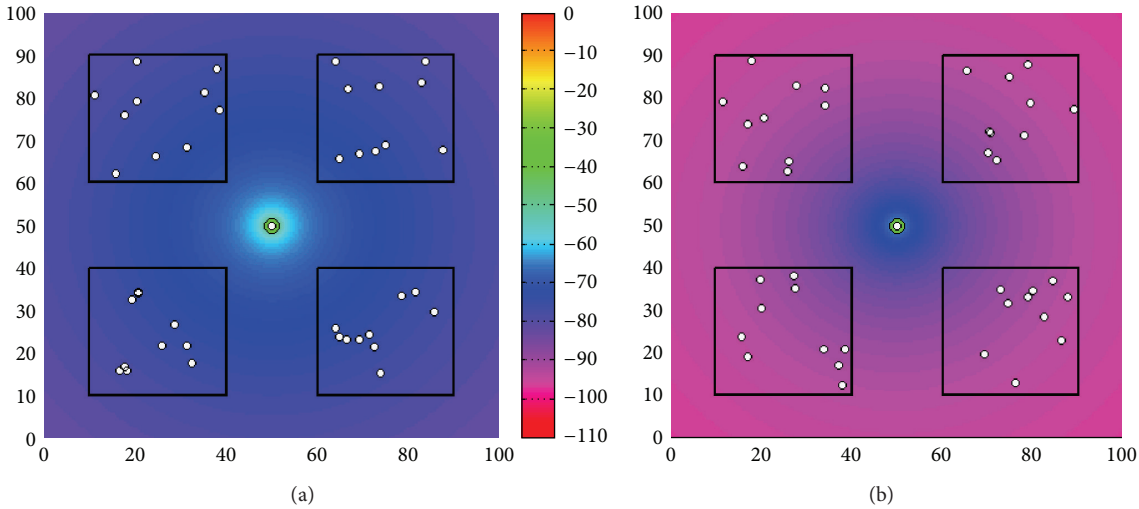


Figure 15: Received signal power, single concentrator in central position: (a) 169 MHz, (b) 868 MHz.

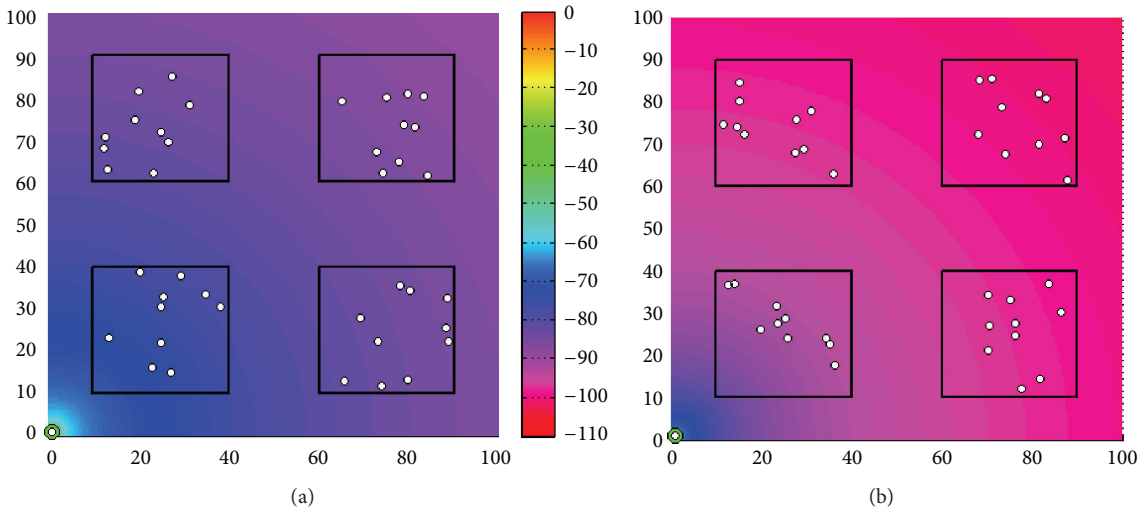


Figure 16: Received signal power, single concentrator in vertex position: (a) 169 MHz, (b) 868 MHz.

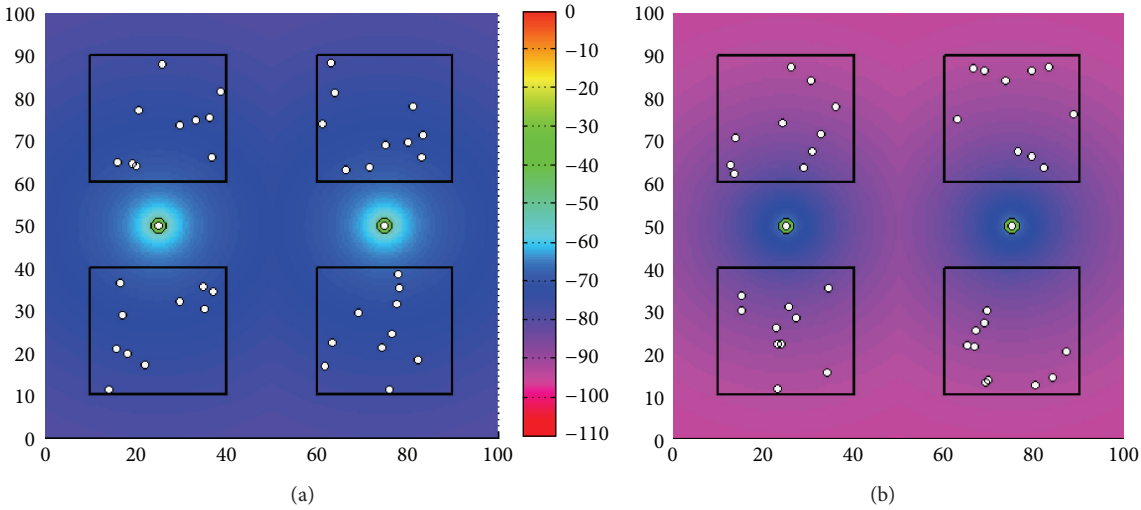


Figure 17: Received signal power, two concentrators in central positions: (a) 169 MHz, (b) 868 MHz.

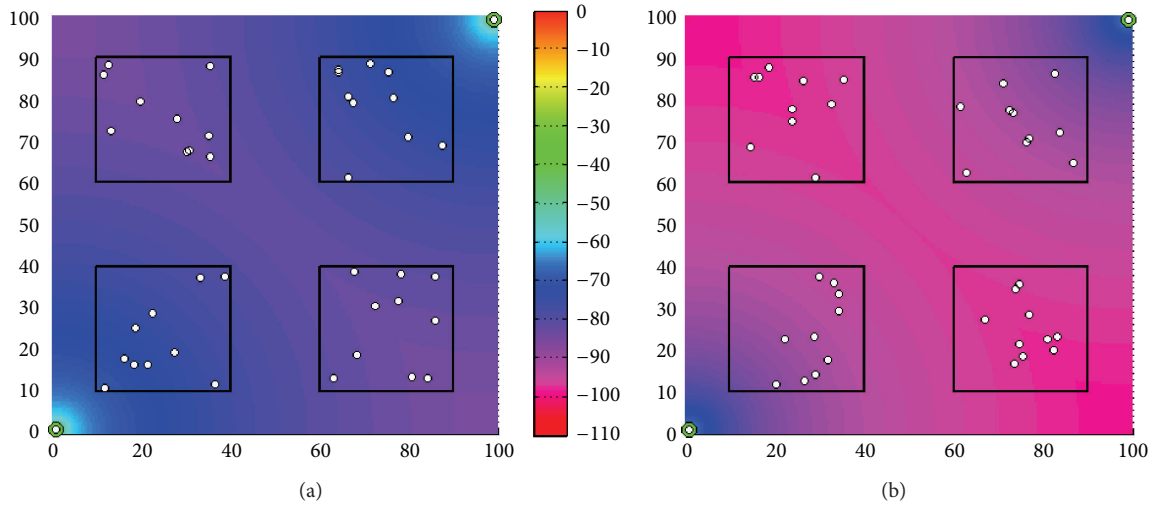


Figure 18: Received signal power, two concentrators in vertex positions: (a) 169 MHz, (b) 868 MHz.

transmit their data independently from each other, without a predetermined channel access mechanism. Each transmission has a time duration denoted by t_p, whereas T represents the time interval between consecutive transmissions of the same node. A collision takes place when one or more sensor nodes start their transmissions within a time interval t_p from the beginning of another node transmission, which prevents the concentrator from correctly receiving the transmitted information.

A network in which nodes may act independently and asynchronously can be modeled as a Poisson process, that is, a stochastic process in which events may take place at any time, independently from each other. Such a process is defined by a random variable N_t, which accounts for the number of events that occurred in the time interval $[0, t_p]$, where $t_p > 0$. The number of events that occur in the time instants from s to t (where $t > s > 0$) is given by $N_t - N_s = j$, with $j = 0, 1, 2, \ldots$, and exhibits a Poisson distribution given by

$$P\{N_t - N_s = j\} = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^j}{j!}. \tag{8}$$

In the equation, λ represents the average of the distribution, that is, the average number of times in which the random event occurs in the interval. This Poisson model is applied to the WM-Bus based WSN. The average time between transmissions from two sensors is T/n. The Poisson process has a rate $\lambda = n/T$, and (8) becomes

$$P\{N_t = j\} = p(j, t_p) = e^{-n(t_p/T)} \frac{\left[n\left(t_p/T\right)\right]^j}{j!}, \tag{9}$$

which represents the probability that the number of sensor nodes transmitting in the time interval $[0, t_p]$ equals j. To compute the probability of collisions in a given time interval, denoted by $P(A_C)$, the following equation is applied:

$$P(A_C) = \sum_{j=2}^{\infty} p(j, t_p) = 1 - p(0, t_p) - p(1, t_p) \tag{10}$$

$$= 1 - e^{-n(t_p/T)} - n \frac{t_p}{T} e^{-n(t_p/T)}. \tag{11}$$

Evaluations of the collisions probability have been performed, according to (10). In a first simulation, a bidirectional communication is assumed, in which the transmission of a WM-Bus packet requires 125 ms (corresponding to N mode at 38.4 kbps). Figure 19 shows the results obtained by varying the time interval between consecutive transmissions. Better performance, that is, a reduced collision probability, is attained when considering a unidirectional communication, as shown in Figure 20. Each sensor node remains in the transmitting state for a short time of 25 ms. For a time interval of 6 hours between consecutive transmissions, which is a quite good value for metering applications, and a concentrator that handles the transmissions of 100 sensor nodes, the collision probability amounts to $1.8 \cdot 10^{-8}$ and $7 \cdot 10^{-9}$, respectively, in the case of bidirectional and unidirectional communications. It is possible to state that, in both cases of bidirectional and unidirectional communications, the WM-Bus WSN is acceptably robust against collisions among the transmissions generated by different sensor nodes. The collision probability, though very small, is not zero; in order to further increase the robustness of the network against collisions, it could be possible to implement an integrity check mechanism on each packet, to possibly request its retransmission by sending a negative acknowledgment. Even more advanced strategies can be conceived to face a possible increase in packet collisions rate, such as the possibility of recovering data from multiple erroneous packets on a stationary receiver, as presented in [31]. Otherwise, a suitable channel access strategy could be designed and implemented.

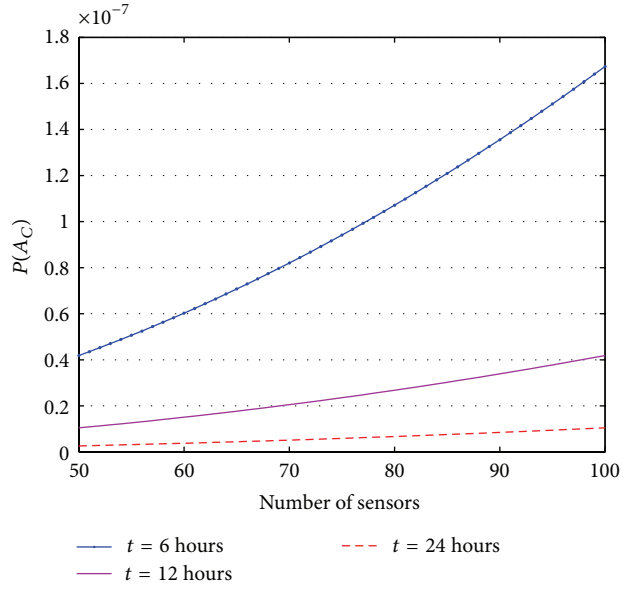


Figure 19: Collision probability as a function of the number of sensors in a bidirectional WSN, for $T = 6, 12, 24$ hours.

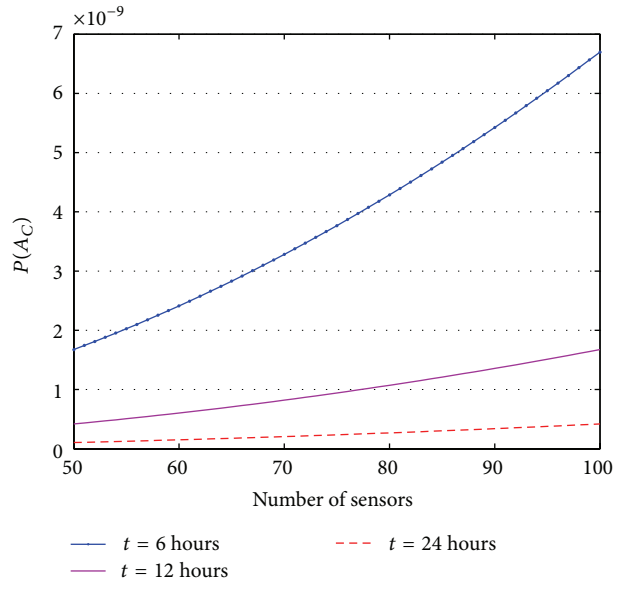


Figure 20: Collision probability as a function of the number of sensors in a unidirectional WSN, for $T = 6, 12, 24$ hours.

6. Final Remarks and Conclusion

This paper investigated the WM-Bus protocol for possible adoption in future smart water grids, by evaluating energy consumption issues and transmission performance in different modes, through simulations and experimental tests over prototype implementations, either hardware or software. The addressed topic is surely one of the hottest in advanced metering infrastructure research area, especially in the European scenario, as confirmed in the last few years by the long debate in which technology options can be regarded as the most viable choice for reliable and efficient capillary network solutions. In authors' opinion, the current work thus represents a useful reference for design purposes and strengthens the research along the main conclusive issues raised in the paper and supported by extensive experimental tests.

The WM-Bus protocol at 169 MHz seems able to ensure adequate transmission capabilities, in different environments, by providing a coverage range in the order of a few kilometers and a quite low sensitivity of the receiver. Bandwidth availability issues, and possible bottlenecks if the number of nodes within the smart water grid increases, shall be put into the perspective of the WM-Bus protocol, which is based on the concept of sporadic data transmission, at a very low duty cycle (as requested in order to freely access the 169 MHz bandwidth for short range radio communications). Should traffic issues arise within the channel, suitable access mechanisms could be included in the framework of the communication policies, thus giving space for a further detailed comparative analysis, to evaluate the most suitable solution.

Experimental tests have shown how the WM-Bus modes at 169 MHz are more demanding in terms of required energy resources, with respect to the 868 MHz counterpart, but are surely fully sustainable, in smart water metering scenarios. In the light of the better coverage properties attainable with a 169 MHz-based communication and considering the importance of such feature, specially when meter locations are severely constrained, the authors are allowed to conclude that the WM-Bus N modes surely represent the best tradeoff for the applicative context under study. Even better outcomes are expected, when the prototype nodes will feature an optimized design and firmware implementation of the standard.

Future efforts will be targeted to the integration of energy-aware task scheduler [32–34] to optimally manage the tasks execution in sensor node processors, the employment of adaptive energy harvesting solutions to improve the node sustainability, and the development of full WSNs based on WM-Bus sensors, for metering scenarios first, and water leakages monitoring then.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this article.

References

[1] S. Lan, M. Qilong, and J. Du, "Architecture of wireless sensor networks for environmental monitoring," in *Proceedings of the International Workshop on Education, Technology, and Training and International Workshop on Geoscience and Remote Sensing (ETT and GRS '08)*, pp. 579–582, Shanghai, China, December 2008.

[2] H. Yuan, L. Yunhao, S. Xingfa, L. Mo, and G. Dai, "Noninteractive localization of wireless camera sensors with mobile beacon," *IEEE Transactions on Mobile Computing*, vol. 12, no. 2, pp. 333–345, 2013.

[3] F. Viani, F. Robol, A. Polo, P. Rocca, G. Oliveri, and A. Massa, "Wireless architectures for heterogeneous sensing in smart home applications: concepts and real implementation," *Proceedings of the IEEE*, vol. 101, no. 11, pp. 2381–2396, 2013.

[4] M. Nicolini, "Optimal pressure management in water networks: increased efficiency and reduced energy costs," in *Proceedings of the Defense Science Research Conference and Expo (DSR '11)*, pp. 1–4, Singapore, August 2011.

[5] G. Sanz, R. Perez, and A. Escobet, "Leakage localization in water networks using fuzzy logic," in *Proceedings of the 20th Mediterranean Conference on Control & Automation (MED '12)*, pp. 646–651, Barcelona, Spain, July 2012.

[6] M. Mencarelli, M. Pizzichini, L. Gabrielli, and S. Squartini, "Self-powered sensor networks for water grids: challenges and preliminary evaluations," *Journal of Selected Areas in Telecommunications*, 2012.

[7] S. Spinsante, M. Pizzichini, M. Mencarelli, S. Squartini, and E. Gambi, "Evaluation of the Wireless M-Bus standard for future smart water grids," in *Proceedings of the 9th International Wireless Communications and Mobile Computing Conference (IWCMC '13)*, pp. 1382–1387, Sardinia, Italy, July 2013.

[8] S. Squartini, L. Gabrielli, M. Mencarelli, M. Pizzichini, S. Spinsante, and F. Piazza, "Wireless M-Bus sensor nodes in smart water grids: the energy issue," in *Proceedings of the 4th International Conference on Intelligent Control and Information Processing (ICICIP '13)*, pp. 614–619, Beijing, China, June 2013.

[9] "Communication systems for meters and remote reading of meters. Part 4: Wireless meter readout (Radio meter reading for operation in the 868 MHz to 870 MHz SRD band)," EN, 13757-4:2005 and EN, 13757-4:2011.

[10] "ZigBee Smart Energy Profile 2.0.," http://www.zigbee.org./.

[11] J. Hui, D. Culler, and S. Chakrabarti, *6LoWPAN: Incorporating IEEE 802.15.4 into the IP Architecture*, White paper # 3, Internet Protocol for Smart Objects (IPSO) Alliance, 2009.

[12] M. Mi, W. Wy, and M. Moniri, "Power harvesting for smart sensor networks in monitoring water distribution system," in *Proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC '11)*, pp. 393–398, Delft, The Netherlands, April 2011.

[13] J. Hayes, K. T. Lau, and D. Diamond, "A wireless sensor network for monitoring water treatment," in *Proceedings of the International Conference on Sensor Technologies and Applications (SENSORCOMM '07)*, pp. 514–519, Valencia, Spain, October 2007.

[14] M. Lin, Y. Wu, and I. Wassell, "Wireless sensor network: water distribution monitoring system," in *Proceedings of the IEEE Radio and Wireless Symposium (RWS '08)*, pp. 775–778, Orlando, Fla, USA, January 2008.

[15] D. C. McKinney and M. D. Lin, "Genetic algorithm solution of groundwater management models," *Water Resources Research*, vol. 30, no. 6, pp. 1897–1906, 1994.

[16] X. Cai, D. C. McKinney, and L. S. Lasdon, "Solving nonlinear water management models using a combined genetic algorithm and linear programming approach," *Advances in Water Resources*, vol. 24, no. 6, pp. 667–676, 2001.

[17] S. Srirangarajan, M. Allen, A. Preis, M. Iqbal, H. Lim, and A. Whittle, "Water main burst event detection and localization," in *Proceedings of the 12th Annual International Conference on Water Distribution Systems Analysis (WDSA '10)*, pp. 1324–1335, Tusan, Ariz, USA, September 2010.

[18] M. Ahadi and M. S. Bakhtiar, "Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to Acoustic Emission signals," *Applied Acoustics*, vol. 71, no. 7, pp. 634–639, 2010.

[19] 2030 Water Resources Group, "Charting Our Water Future. Economic frameworks to inform decision-making," 2009, http://www.mckinsey.com/Client_Service/Sustainability/Latest_thinking/Charting_our_water_future/.

[20] World Bank, *The 2012 World Bank Annual Report*, http://siteresources.worldbank.org/EXTANNREP2012/Resources/8784408-1346247445238/AnnualReport2012En.pdf/.

[21] The OMS Group, *Open Metering Systems*, http://www.oms-group.org/en index.html/.

[22] T. S. Rappaport, *Wireless Communication*, Prentice Hall, New York, NY, USA, 1996.

[23] "Communication systems for meters and remote reading of meters. Part 5: Wireless Relaying," EN, 13757-5, 2009.

[24] Texas Instruments, *Smart RF Transceiver Evaluation Board—TrxEB User's Guide*, 2012, http://www.ti.com/general/docs/lit/getliterature.tsp?baseLiteratureNumber=swru294.

[25] Texas Instruments, *MSP430x2xx Family User'S Guide*, 2011, http://www.element14.com/community/servlet/JiveServlet/previewBody/40110-102-1-225744/TexasInstruments.UsernGuide-n1.pdf.

[26] Texas Instruments, *CC1120 Data Sheet*, 2012, http://www.ti.com/lit/ds/symlink/cc1120.pdf.

[27] P. Seem, *Wireless MBUS Implementation with CC1101 and MSP430*, Texas Instruments Application Note 067.

[28] F. Di Franco, C. Tachtatzis, B. Graham et al., "Current characterisation for ultra low power wireless body area networks," in *Proceedings of the 8th IEEE Workshop on Intelligent Solutions in Embedded Systems (WISES '10)*, pp. 91–96, Heraklion, Greece, July 2010.

[29] Texas Instruments, *CC112X/CC1175 Low-Power High Performance Sub-1 GHz RF Transceivers/Transmitter*, 2011.

[30] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, 1980.

[31] R. M. Jacobsen and P. Popovski, "Data recovery using side information from the wireless M-Bus protocol," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP '13)*, pp. 511–514, Austin, Tex, USA, December 2013.

[32] M. Severini, S. Squartini, and F. Piazza, "Energy-aware lazy scheduling algorithm for energy-harvesting sensor nodes," *Neural Computing and Applications*, vol. 23, no. 7-8, pp. 1899–1908, 2013.

[33] H. El Ghor, M. Chetto, and R. H. Chehade, "A nonclairvoyant real-time scheduler for ambient energy harvesting sensors," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 732652, 11 pages, 2013.

[34] A. R. Silva, M. Liu, and M. Moghaddam, "An adaptive energy-management framework for sensor nodes with constrained energy scavenging profiles," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 272849, 33 pages, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 627689, 18 pages
http://dx.doi.org/10.1155/2014/627689

Research Article

ZEQoS: A New Energy and QoS-Aware Routing Protocol for Communication of Sensor Devices in Healthcare System

Zahoor Ali Khan,[1] Shyamala Sivakumar,[2] William Phillips,[1] and Bill Robertson[1]

[1] *Internetworking Program, Faculty of Engineering, Dalhousie University, Halifax, NS, Canada B3H 4R2*
[2] *Saint Mary's University, Halifax, NS, Canada B3H 3C3*

Correspondence should be addressed to Zahoor Ali Khan; zahoor.khan@dal.ca

Received 2 November 2013; Revised 18 February 2014; Accepted 19 February 2014; Published 5 June 2014

Academic Editor: Christos Verikoukis

This paper proposes a novel integrated energy and QoS-aware routing protocol with the considerations of energy, end-to-end latency, and reliability requirements of body area network (BAN) communication. The proposed routing protocol, called ZEQoS, introduces two main modules (MAC layer and network layer) and three algorithms (neighbor table constructor, routing table constructor, and path selector). To handle ordinary packets (OPs), delay-sensitive packets (DSPs), and reliability-sensitive packets (RSPs), the new mechanism first calculates the communication costs, end-to-end path delays, and end-to-end path reliabilities of all possible paths from a source to destination. The protocol then selects the best possible path(s) for OPs, RSPs, and DSPs by considering their QoS requirement. Extensive simulations using OMNeT++ based simulator Castalia 3.2 demonstrate that the performance of the proposed integrated algorithm is satisfactory when tested on a real hospital scenario, and all data types including OPs, DSPs, and RSPs are used as offered traffic. Simulations also show that the ZEQoS also offers better performance in terms of higher throughput, less packets dropped on MAC and network layers, and lower network traffic than comparable protocols including DMQoS and noRouting.

1. Introduction

Various advanced and valuable state-of-the-art applications of body area networks (BANs) help enhance the patient's healthcare monitoring and their quality of life. The BAN devices are used to monitor the patients' health related concerns such as changes in blood pressure (BP), heart rate, or body temperature. In BAN communication, the body implanted and wearable sensors send their data to a central and computationally more powerful device known as the coordinator. The coordinator also behaves like a router in BAN networks. The BAN sensor nodes are, typically, required to use an extremely low transmission power to reduce health concerns and avoid tissue heating [1]. The low transmit power restricts the BAN transmission range to few meters (approximately three meters) [2]. One of the BAN features is to facilitate the physical mobility of the patient; this means that now the patients are not required to stay in the hospital at all times. Routing protocols are required to route a patient's data towards the required destination even when a patient moves. Routing is an issue for the sensor nodes due to the limited availability of resources including ultralow computation power, lower memory, and reduced energy source. The radio frequency (RF) portion of the sensor nodes in BAN plays a major role in the consumption of energy. MAC protocols can reduce the energy consumption by controlling the duty cycle of the RF part. MAC protocols are also helpful in effectively controlling the other sources that are the cause of energy waste, such as collision, idle listening, overhearing, and packet overhead. In short, an ideal MAC protocol increases error-free data transmission, maximum throughput, and medium access management and minimizes transmission delay, thereby increasing network lifetime. Despite the fact that MAC protocols are helpful in resolving many problems, the issues of end-to-end packet delivery, logical-physical address mapping, frame fragmentation, addressing techniques, and route determination methods are not in the scope of MAC protocols. These issues can

be more easily handled by the network layer. As a result, it is important to consider the network layer routing protocols to resolve these issues [3].

The challenges and features of BAN are different than WSN due to the specific needs of the wireless environment on the human body. The development of an efficient routing protocol in BAN requires more careful considerations than WSN. Some of the important factors to consider for the BAN routing protocols are their limited bandwidth, node and link heterogeneity, energy efficiency, coverage area, data aggregation, quality of service (QoS), transmit power, and mobile flexibility [3, 4].

The effect of fading, noise, and interference plays important role to reduce the effective bandwidth. The bandwidth available for BAN also varies due to these effects. The routing protocol can have only limited network control. The placement of sensor nodes during the formation of BAN is possible by a manual process. The nodes are placed manually on the predefined locations of the body where the data transmission is minimally disturbed by noise or interference. Ideally each node sends its own data and forwards the data received from other nodes towards the required destination. But in case of BAN, the implanted sensors, due to their tiny size and limited energy resources, only send the data to the central node or coordinator. The coordinator and other wearable nodes are capable of multihop communication, which helps route the data towards the desired destination. With the consideration of these facts, the routing protocol should be able to find and manage alternate routing paths in case of node failure.

Most of the nodes used in BAN are heterogeneous in terms of their capabilities including available energy, computational power, and communication capability. An example of heterogeneous nodes in BAN is the use of different wearable sensors to monitor body temperature, blood pressure, and other important vital signs of a patient. The link speeds of different implanted and wearable sensor nodes are not similar. The heterogeneity of the nodes should be considered by routing protocols.

The sensor nodes are placed on a human body that can be in motion. The node functionality may be affected due to mobility of the patient. This is because, the sensing capability of the mobile node can place increased energy demands on an application in different scenarios, for example, vital sign monitoring of a mobile patient indoor in the hospital is different than a patient in the outside environment of the hospital. With the mobility of the nodes, the routing protocol should be able to provide a suitable solution for the reliable communication.

Quality of service is one of the important factors in BAN communication. The reliability of associated algorithms improves the successful delivery of critical reliability-sensitive data from sensor nodes to the base station. The routing protocols fulfill the QoS demand of different BAN applications by using the delay-control algorithms. These QoS-aware protocols help monitor the patient's health during a critical situation [5, 6]. Our proposed routing protocol based on ordinary, delay-sensitive, and reliability-sensitive data is for the indoor hospital environment with the enhanced capability of handling mobile node communications.

The paper is organized as follows. Section 2 provides the motivation of this protocol. Section 3 explains the proposed routing protocol (ZEQoS). Sections 4 and 5 provide the MAC and network layer modules, respectively. Section 6 discusses the performance evaluation of ZEQoS. Section 7 demonstrates the superior performance of ZEQoS when compared with DMQoS and noRouting, and Section 8 summarizes this paper.

2. Related Work and Motivation

The consideration of quality of service (QoS) is an important but challenging task for the designers of BAN routing protocols. An ideal BAN routing protocol should provide an efficient and reliable path to route the patient's ordinary and critical data. The two important QoS routing protocols are reliability and delay-tolerant based protocols. The reliability-aware routing protocols ensure the delivery of maximum data packets to the destination. The transmission delay is not an issue for the reliability packets' delivery. For achieving the maximum throughput, data packets are sent on multiple redundant paths in some of the techniques used in reliability-aware protocols.

The delay-tolerant based routing protocols deal with the packets that are required to be delivered within a deadline. The route determination for the traffic of video streaming is one of the examples of this kind of routing. The end-to-end packet delay must be less than a specific delay; otherwise, the quality of overall data monitoring will be affected. Many routing protocols are proposed by researchers to address this issue. Researchers have proposed different energy and QoS-aware based routing protocols [7–16]. Some of the important QoS-aware routing protocols such as QoS-aware framework [9], RL-QRP [11], LOCALMOR [13], and DMQoS [17] are briefly discussed below.

In [9], a QoS-aware routing service framework for biomedical sensor networks is proposed based on a cross layered modular approach. The metrics considered for the determination of routes are wireless channel status, packet priority level, and sensor node's willingness to behave as a router. The proposed framework contains four main modules: an application programming interfaces (APIs) module, a routing service module, a packet queuing and scheduling module, and a system information repository module. The APIs module works as an interface between the user application and the routing service module. The components of APIs are QoS metrics selection, packet sending/receiving, packet priority level setting, and admission control and service level control. The QoS metrics are end-to-end delay, delivery ratio, and power consumption. The sensed data sent by user application for sink or other nodes is received by the packet sending/receiving component of APIs. These data packets contain destination ID, source ID, priority level, and payload. The data packets are received from the network layer. The payloads are forwarded to the user application for aggregation after separation from the data packets. The QoS-aware framework [9] is based on a modular technique that addresses QoS related issues for BAN. The newer routing techniques that consider the geographic location of neighbor

nodes prove very effective. The benefits of using geographic based routing include scalability, routing decisions based on neighborhood information, and being adaptive to dynamic environments. These protocols are also effective for mobile nodes. In this paper, the proposed protocol uses a similar modular approach but with the additional enhancements of location and energy aware routing.

RL-QRP [11] is a reinforcement learning based routing protocol with QoS support for biomedical sensor networks. The protocol focuses on two types of QoS requirements: packet delivery ratio and end-to-end delay. The machine learning approach used in this protocol uses optimal routing policies. These optimal routing policies can be found through experiences and rewards without the requirement of keeping precise network state information. RL-QRP [11] considers the neighborhood node's Q-values and location information for the determination of a QoS route. Energy is one of the major constraints in sensor nodes. The drawback of RL-QRP [11] is not considering energy at all. The proposed routing protocol, in this paper, considers the residual energy and geographic location of the next hop node, which helps improve the node lifetime.

LOCALMOR [13] is a QoS based BAN routing protocol that relies on the traffic diversity of biomedical applications and guarantees differentiated routing, based on using QoS metrics. The three different QoS requirements: (1) energy efficiency, (2) reliability, and (3) latency are considered in this protocol. The data traffic of biomedical applications is divided into four classes: regular, reliability-sensitive, delay-sensitive, and critical. A modular approach used in LOCALMOR consists of four modules: a power-efficiency module, a reliability-sensitive module, a delay-sensitive module, and a neighbor manager. Hello packets are used to update the neighbor's information in the neighbor table. The neighbor manager module is responsible to send/receive the Hello packets and manage the update of information of neighbors. The data from body sensor nodes transfer to the primary and secondary sinks via routers. LOCALMOR [13] provides a QoS-aware modular solution for different packet types. A data-centric multiobjective QoS-aware routing protocol (DMQoS) [17] outperforms the LOCALMOR [13]. The modular based architecture of DMQoS [17] provides the different routing modules to fulfill the QoS services for different packet classes. The reliability and delay control modules introduced in [17] result in better performance than several state-of-the-art approaches [11, 12, 15, 18–22] in terms of lower bit error rates, traffic load, and operation energy overload. The purpose of proposed energy and QoS-aware routing protocol (ZEQoS) is the reliable and energy-efficient routing similar to LOCALMOR and DMQoS. In this paper, the proposed routing protocol uses a similar modular approach and same packet classification as discussed in LOCALMOR and DMQoS. However, the mechanism of Hello protocol and calculation used for end-to-end path delays and end-to-end path reliabilities improves throughput and reduces the network traffic load. The simulation results prove that our protocol, ZEQoS, performs better than these protocols.

An energy-aware peering routing protocol (EPR) discussed in [23] is used to choose the best next hop for only ordinary packets (OPs) by considering the energy availability and geographic information of the devices. EPR has an overall lower energy consumption than comparable protocols [12, 15, 17, 20, 21] and provides better results in terms of reduced overall network traffic, reduced number of packets forwarded by intermediate nodes, and higher successful data transmission rates. In [5], the QPRD was extended to consider delay-sensitive packets (DSPs) as well as OPs. The resulting QPRD proposed an algorithm to route DSPs in addition to OPs. The redundant paths with the help of end-to-end path reliabilities are used in QPRR, discussed in [6], to ensure the reliable transmission of reliability-sensitive packets (RSPs) and OPs. These proposed routing protocols EPR, QPRD, and QPRR are not capable of handling OPs, RSPs, and DSPs simultaneously. For real-time display of patient data in the hospital environment, an energy and QoS-aware routing protocol is required that can handle all three data types (i.e., OPs, DSPs, and RSPs) simultaneously. With the integration of EPR, QPRD, and QPRR, in a unified BAN routing protocol, the ZEQoS provides a reliable solution for the transmission of OPs, RSPs, and DSPs and displays real-time BAN data.

3. Proposed Energy and QoS-Aware Routing Protocol (ZEQoS)

The proposed Zahoor energy and QoS-aware routing protocol (ZEQoS) is intended to be associated with the indoor hospital ZK-BAN peering framework [23]. To summarize, the ZK-BAN peering framework categorizes hospital devices into three types with the consideration of their energy levels. Figure 1 shows a general BAN communication framework. This hierarchical model has three communication tiers [24]. The sensor devices connected to body send data to the BAN coordinator (BANC) in tier 1. The BANC behaves like a cluster-head in WSNs. In tier 2, the possible next hop of a BANC is a BANC, medical display coordinator, nursing station coordinator, or a cellular device as shown in Figure 1. The tier 2 communication devices with the exclusion of the BANC forward the BAN data to tier 3 communication devices.

The device directly connected with the power source is considered as type 1 device such as nursing station coordinator (NSC). Devices with replaceable batteries (e.g., medical display coordinators (MDCs)) and nonreplaceable batteries (e.g., body area network coordinators (BANCs)) are counted in type 2 and type 3 devices, respectively, and this is illustrated in Table 1.

According to ZK-BAN peering framework [23], the information of BANCs and their respective peer MDCs are stored at the NSC. This framework uses a hybrid communication mode that can be in one of two modes, a centralized mode or a distributed mode as appropriate. Hybrid communication helps increase privacy and save energy consumption. In centralized mode, the BANCs get the information of its respective peer from the NSC. In distributed mode, BANCs send the data reliably to their peer MDC in order to achieve the purpose of real-time display of patient data. The detailed

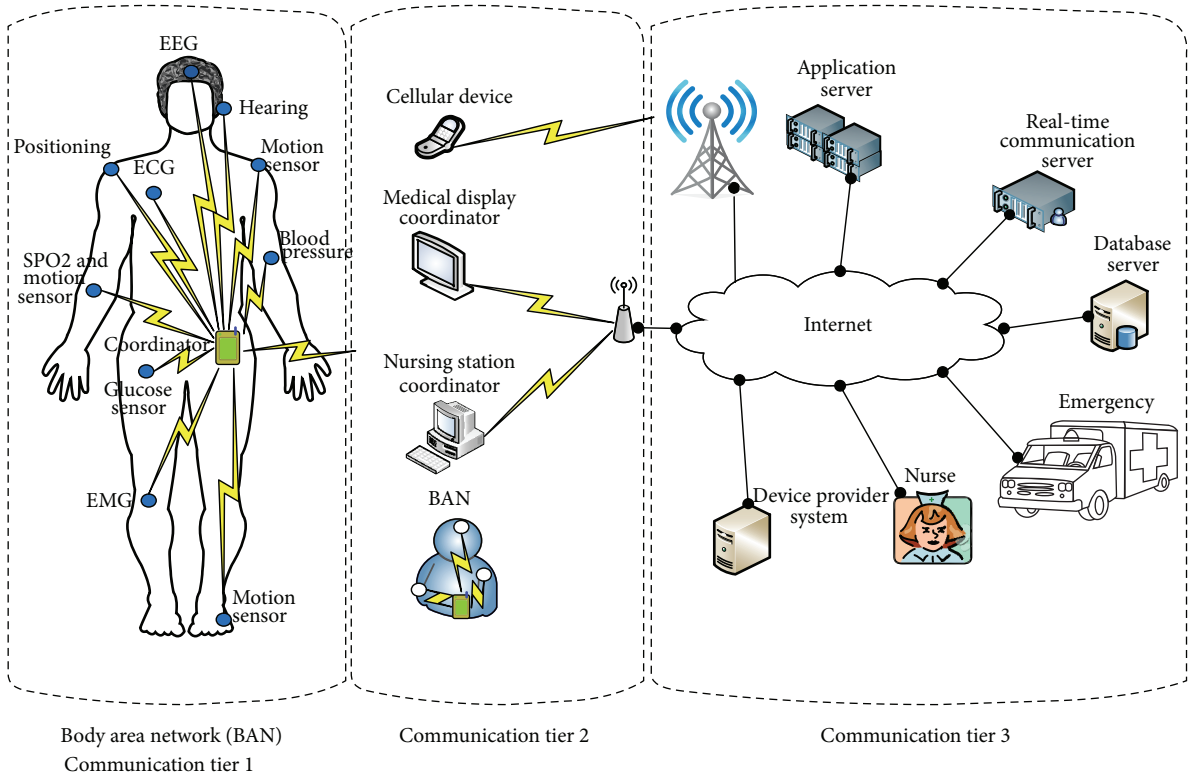


Figure 1: General BAN communication system.

Table 1: Classification of devices in hospital environment.

Class	Device name	Power source	Channels	MAC protocol	Mobility
1	NSC	Directly connected	2	IEEE 802.15.4 IEEE 802.11	No
2	MDC	Replaceable batteries	2	IEEE 802.15.4 IEEE 802.11	Yes
3	BANC	Limited energy available	1	IEEE 802.15.4	Yes

discussion of ZK-BAN peering framework can be found in [23].

ZEQoS calculates the best next hops for OPs, DSPs, and RSPs with the help of different modules and algorithms. The next hop for OPs is denoted by NH_E. The selection of NH_E is based on the communication cost (C_i) which is calculated with the consideration of geographic and energy information of the neighbor nodes. The ZEQoS employs a Hello protocol, discussed in [23], that is used to broadcast the important information of a node to the other nodes. For DSPs, ZEQoS calculates the node delay and end-to-end path delays of all possible paths from source to destination and then chooses the next hop (i.e., NH_D) device based on the lowest end-to-end path delay. For RSPs, ZEQoS (1) computes the end-to-end path reliabilities of all possible paths, (2) selects the three most reliable paths for each destination, (3) determines the degree of path redundancy, and (4) chooses the next hop device(s) based on the most reliable end-to-end path(s) from the source node to the destination. ZEQoS improves the reliability with the help of redundant paths. The architecture of proposed ZEQoS routing protocol is shown in Figure 2 and notations used in this protocol are given in Table 2.

The modules used in ZEQoS are spread into two layers: MAC layer and network layer. MAC and network layer modules are discussed below.

4. MAC Layer Modules

The MAC layer contains four modules: MAC receiver, reliability module, delay module, and MAC transmitter. The data or Hello packets from other nodes (i.e., BANC, MDC, or NSC) are received by MAC receiver of the node i. MAC receiver checks the MAC address of the packets and only forwards the packets, which contain the broadcast address or MAC address of the node i as destination address, to the network layer. The reliability module of node i on MAC layer calculates the numbers of packets sent to neighbor node j and the number of acknowledgements received

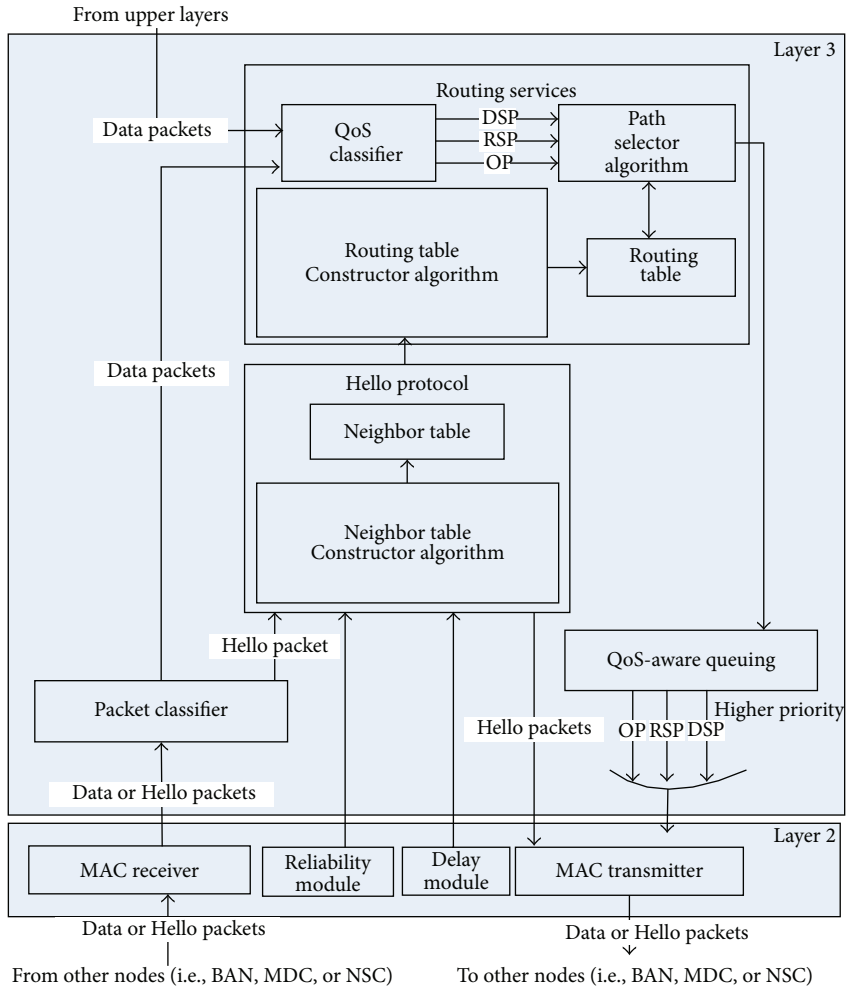


Figure 2: ZEQoS routing protocol architecture.

ID_{Dst}	L_{Dst}	ID_j	L_j	$D_{(j,Dst)}$	E_j	T_j	$R_{path(j,Dst)}$	$DL_{path(j,Dst)}$

Figure 3: Hello packet structure.

from neighbor node j. The delay module monitors the time required to capture the channel ($DL_{channel(i)}$), MAC layer queuing delay ($DL_{MAC\_queue(i)}$), and transmission time ($DL_{trans(i)}$) of a packet. The delay and reliability modules send their information to the Hello protocol module of the network layer. The neighbor table constructor algorithm in Hello protocols module uses this information to calculate the node delay ($DL_{node(i)}$) and the link reliability between the node i and the neighbor node j ($R_{link(i,j)}$).

The data and Hello packets from the network layer are received by the MAC transmitter submodule which stores these packets in the MAC layer queue. The MAC layer queue works in a first-in-first-out (FIFO) fashion. MAC transmitter uses CSMA/CA algorithm to send the data when the channel is captured.

5. Network Layer Modules

Network layer consists of four modules: packet classifier (PC), Hello protocol module (HPM), routing services module (RSM), and QoS-aware queuing module (QQM). The detailed discussion of these modules is given below.

5.1. Packet Classifier. The packet classifier receives data and Hello packets from the MAC receiver module of the MAC layer. The job of packet classifier is to differentiate and forward the data packets and Hello packets to the routing services module and Hello protocol module, respectively.

5.2. Hello Protocol Module (HPM). According to the Hello protocol, type 1 and type 2 devices (NSC or MDCs) send Hello packets periodically and the BANCs broadcast their Hello packets only at the reception of other nodes' Hello packets which contain the NSC or MDC information. The Hello packet fields of node j are shown in Figure 3. The possible destination (Dst) can be a NSC, MDC, or BANC. The Hello packet contains the information about the destination device

Table 2: Notations for the proposed algorithm.

Field ID	Description
Node i	Source node
Node j	Neighbor node of source node
Node Dst.	Destination node (i.e. NSC, MDCs, and BAN)
$\mathrm{ID}_{\mathrm{Dst}}$	Destination ID
L_{Dst}	Destination location
ID_j	Neighbor node j ID
L_j	Neighbor node j location
$D_{(j,\mathrm{Dst})}$	Distance between neighbor node j and destination Dst.
E_j	Residual energy of node j
C_j	Communication cost
T_j	Device type of node j
$R_{\mathrm{path}(j,\mathrm{Dst})}$	Path reliability between neighbor j and destination
$D_{(i,j)}$	Distance between node i to neighbor node j
$R_{\mathrm{link}(i,j)}$	Link reliability from node i to neighbor node j
$R_{\mathrm{path}(i,\mathrm{Dst})}$	Path reliability from node i to destination Dst.
$\mathrm{NH}_{(i,\mathrm{Dst})}$	Next hop between node i and destination Dst.
NH_E	Energy-aware next hop
NH_{R1}	1st reliable next hop
NH_{R2}	2nd reliable next hop
NH_{R3}	3rd reliable next hop
NH_D	Next hop for delay-sensitive packets
$\mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})}$	Path delay from node i to destination Dst.
$\mathrm{DL}_{\mathrm{node}(i)}$	Time delay within the node i
$\mathrm{DL}_{\mathrm{req}}$	Required path delay for delay-sensitive packets
R_{req}	Required reliability of reliability-sensitive packets
$R_{\mathrm{option1}(i,\mathrm{Dst})}$	1st option reliability for sending reliability-sensitive packets
$R_{\mathrm{option2}(i,\mathrm{Dst})}$	2nd option reliability for sending reliability-sensitive packets
$R_{\mathrm{option3}(i,\mathrm{Dst})}$	3rd option reliability for sending reliability-sensitive packets

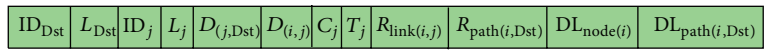


Figure 4: Neighbor table structure.

ID ($\mathrm{ID}_{\mathrm{Dst}}$), destination location (L_{Dst}), sender's ID (ID_j), residual energy (E_j), device type (T_j), distance ($D_{(j,\mathrm{Dst})}$), path reliability ($R_{\mathrm{path}(j,\mathrm{Dst})}$), and path delay ($\mathrm{DL}_{\mathrm{path}(j,\mathrm{Dst})}$). The subscript (j, Dst) means from sender node j to the destination.

The node i receives the Hello packet. The information received from the reliability module, delay module, and Hello packets of the MAC receiver module is used by the neighbor table constructor algorithm to construct the neighbor table. The neighbor table constructor algorithm of node i calculates its own $\mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})}$ and $R_{\mathrm{path}(i,\mathrm{Dst})}$ based on the information in the Hello packets. Node i updates the values of Hello packet fields and broadcasts it to the other nodes. The mechanism of Hello protocol used in ZEQoS is described briefly in the following paragraph and is described in [23].

The neighbor table and neighbor table constructor algorithm are the two submodules of the Hello protocol module. In addition to Hello packet fields, the neighbor table contains fields for both hop-by-hop delay ($\mathrm{DL}_{\mathrm{node}(i)}$) and reliability ($R_{\mathrm{link}(i)}$) and end-to-end delay ($\mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})}$) and reliability ($R_{\mathrm{path}(i,\mathrm{Dst})}$). Neighbor table also uses communication cost (C_i) instead of residual energy (E_i). The neighbor table structure of node i is shown in Figure 4.

5.2.1. Neighbor Table Constructor Algorithm. The neighbor table constructor algorithm updates the values of the neighbor table fields periodically after receiving every new Hello packet. Neighbor table constructor algorithm calculates the values of the additional field used in neighbor table such as $\mathrm{DL}_{\mathrm{node}(i)}$, $R_{\mathrm{link}(i)}$, $\mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})}$, $R_{\mathrm{path}(i,\mathrm{Dst})}$, C_i, and $D_{(i,j)}$. The terms rm, hp, dm, and nt used in Algorithm 1 stand for reliability module, Hello packet, delay module, and neighbor table, respectively.

INPUT: Hello Packet, at each node i.

(1) $\overline{X}_i = \dfrac{N_{\text{Acks}}(\text{rm})}{N_{\text{Trans}}(\text{rm})}$
(2) $\rho_r \leftarrow 0.4$
(3) $R_{\text{link}(i,j)} = (1 - \rho_r) * R_{\text{link}(i,j)} + \rho_r * \overline{X}_i$
(4) $R_{\text{path}(i,\text{Dst})} = R_{\text{link}(i,j)} + R_{\text{path}(j,\text{Dst})}(\text{hp})$
(5) $\rho_d \leftarrow 0.2$
(6) $\text{DL}_{\text{queue+channel}} \leftarrow$ First packet delay
(7) $\text{DL}_{\text{queue+channel}} = (1 - \rho_d) * \left(\text{DL}_{\text{MAC\_queue}}(\text{dm}) + \text{DL}_{\text{channel}}(\text{dm}) + \text{DL}_{\text{Net\_queue}}\right) + \rho_d * \left(\text{DL}_{\text{MAC\_queue}}(\text{dm}) + \text{DL}_{\text{channel}}(\text{dm}) + \text{DL}_{\text{Net\_queue}}\right)$
(8) $\text{DL}_{\text{node}(i)} = \text{DL}_{\text{trans}(i)}(\text{dm}) + \text{DL}_{\text{queue+channel}} + \text{DL}_{\text{proc}}$
(9) $\text{DL}_{\text{path}(i,\text{Dst})} = \text{DL}_{\text{node}(i)} + \text{DL}_{\text{path}(j,\text{Dst})}(\text{hp})$
(10) $D_{(i,j)} = \sqrt{\left(X_i - X_j\right)^2 + \left(Y_i - Y_j\right)^2}$
(11) $C_j = \dfrac{\left(T_j(\text{hp}) * D^2_{(i,j)}(\text{hp})\right)}{E_j(\text{hp})}$
(12) $D_{(i,\text{Dst})} = \sqrt{\left(X_i - X_{\text{Dst}}\right)^2 + \left(Y_i - Y_{\text{Dst}}\right)^2}$
(13) **if** $\left(D_{(j,\text{Dst}}(\text{hp}) < D_{(i,\text{Dst})}\right)$ **then**
(14) (add a new record for the Dst's information in the neighbor table)
(15) $\text{ID}_{\text{Dst}}(\text{nt}) \leftarrow \text{ID}_{\text{Dst}}(\text{hp})$
(16) $\text{ID}_j(\text{nt}) \leftarrow \text{ID}_j(\text{hp})$
(17) $L_j(\text{nt}) \leftarrow L_j(\text{hp})$
(18) $D_{(j,\text{Dst})}(\text{nt}) \leftarrow D_{(j,\text{Dst})}(\text{hp})$
(19) $D_{(i,j)}(\text{nt}) \leftarrow D_{(i,j)}$
(20) $C_j(\text{nt}) \leftarrow C_j$
(21) $T_j(\text{nt}) \leftarrow T_j(\text{hp})$
(22) $R_{\text{link}(i,j)}(\text{nt}) \leftarrow R_{\text{link}(i,j)}$
(23) $R_{\text{path}(i,\text{Dst})}(\text{nt}) \leftarrow R_{\text{path}(i,\text{Dst})}$
(24) $\text{DL}_{\text{node}(i)}(\text{nt}) \leftarrow \text{DL}_{\text{node}(i)}$
(25) $\text{DL}_{\text{path}(i,\text{Dst})}(\text{nt}) \leftarrow \text{DL}_{\text{path}(i,\text{Dst})}$
(26) **end if**
(27) (add a new record for the neighbor node j's information in the neighbor table)
(28) $\text{ID}_{\text{Dst}}(\text{nt}) \leftarrow \text{ID}_{(\text{Dst})}(\text{hp})$
(29) $\text{ID}_j(\text{nt}) \leftarrow \text{ID}_j(\text{hp})$
(30) $L_j(\text{nt}) \leftarrow L_j(\text{hp})$
(31) $D_{(j,\text{Dst})}(\text{nt}) = 0$
(32) $D_{(i,j)}(\text{nt}) \leftarrow D_{(i,j)}$
(33) $C_j(\text{nt}) \leftarrow C_j$
(34) $T_j(\text{nt}) \leftarrow T_j(\text{hp})$
(35) $R_{\text{link}(i,j)}(\text{nt}) \leftarrow R_{\text{link}(i,j)}$
(36) $R_{\text{path}(i,\text{Dst})}(\text{nt}) \leftarrow R_{\text{path}(i,\text{Dst})}$
(37) $\text{DL}_{\text{node}(i)}(\text{nt}) \leftarrow \text{DL}_{\text{node}(i)}$
(38) $\text{DL}_{\text{path}(i,\text{Dst})}(\text{nt}) \leftarrow \text{DL}_{\text{path}(i,\text{Dst})}$

Algorithm 1: Neighbor table constructor algorithm for ZEQoS.

The average probability of successful transmission $\overline{X}_i$ after every 4 seconds is calculated by using (1). Consider

$$\overline{X}_i = \frac{N_{\text{Acks}}}{N_{\text{Trans}}}, \tag{1}$$

where N_{Acks} = number of acknowledgements and N_{Trans} = number of transmissions.

The link reliability between node i and neighbor node j ($R_{\text{link}(i,j)}$) is calculated by using the exponentially weighted moving average (EWMA) equation (2). Consider

$$R_{\text{link}(i,j)} = (1 - \rho_r) R_{\text{link}(i,j)} + \rho_r \times \overline{X}_i, \tag{2}$$

where ρ_r is the average weighting factor that satisfies $0 < \rho_r \leq 1$. Algorithm 1 uses $\rho_r = 0.4$.

The path reliability between node i and destination node Dst ($R_{\text{path}(i,\text{Dst})}$) is calculated by using (3). Consider

$$R_{\text{path}(i,\text{Dst})} = R_{\text{link}(i,j)} \times R_{\text{path}(j,\text{Dst})}. \tag{3}$$

ID_{Dst}	L_{Dst}	NH_E	NH_{R1}	NH_{R2}	NH_{R3}	NH_D	$R_{option1(i,Dst)}$	$R_{option2(i,Dst)}$	$R_{option3(i,Dst)}$	$DL_{path(i,Dst)}$

Figure 5: Routing table structure.

The values of $R_{link(i,j)}$ and $R_{path(j,Dst)}$ are used from (2) and Hello packet (hp), respectively. The calculation of finding $R_{path(i,Dst)}$ is given in Algorithm 1 (lines 1–4).

The delay due to the queues of MAC and network layers and channel capture ($DL_{queue+channel}$) is calculated by using the exponentially weighted moving average (EWMA) formula. Consider

$$\begin{aligned} DL_{queue+channel} &= (1-\rho_d) * \left(DL_{MAC\_queue}(dm) + DL_{channel}(dm) + DL_{Net\_queue}\right) \\ &+ \rho_d * \left(DL_{MAC\_queue}(dm) + DL_{channel}(dm) + DL_{Net_{queue}}\right), \end{aligned} \quad (4)$$

where the values of MAC queue delay and channel capture time are received from delay module (dm), whereas the values of network queue delays are calculated on network layer. The initial value of $DL_{queue+channel}$ is the delay of the first packet sent by the node. The selection of ρ_d value is the personal choice and experience, but it should satisfy $0 < \rho_d \leq 1$. The recommended values are $0.2 \leq \rho_d \leq 0.3$. Algorithm 1 uses $\rho_d = 0.2$.

The value of node delay ($DL_{node(i)}$) is calculated with the addition of the packet delays due to transmission, queuing, processing, and capturing of the channel. Consider

$$DL_{node(i)} = DL_{trans(i)}(dm) + DL_{queue+channel} + DL_{proc}. \quad (5)$$

The path delay between node i and destination node Dst ($DL_{path(i,Dst)}$) is calculated by using (6). Consider

$$DL_{path(i,Dst)} = DL_{node(i)} + DL_{path(j,Dst)}(hp), \quad (6)$$

where initial value of $DL_{path(j,Dst)}$ is zero when $j = Dst$.

The values of $DL_{node(i)}$ are calculated in (5) and $DL_{path(n,Dst)}$ is received from Hello packet (hp).

The calculation of finding $DL_{path(i,Dst)}$ is shown in Algorithm 1 from lines 5–9.

Algorithm 1 (lines 11-12) calculates the communication cost (C_j) and distance from node i to the neighbor node $j(D_{(i,j)})$ by using (7). Consider

$$\begin{aligned} D_{(i,j)} &= \sqrt{\left(X_i - X_j\right)^2 + \left(Y_i - Y_j\right)^2}, \\ C_j &= \frac{\left(T_j * D_{(i,j)}^2\right)}{E_j}, \end{aligned} \quad (7)$$

where X_i, Y_i stand for the X, Y coordinates of node i and X_{DST}, Y_{DST} represent the X, Y coordinates of the destination. It is also assumed that the locations of NSC and MDCs are known. The RSSI localization technique given in [25] is used to calculate the values of X_i and Y_i of the node i. The values of T_j, $D_{(i,j)}$, and E_j are received from Hello packet (hp). The shorter distance ($D_{(i,j)}$), lower device type (T_j), and higher residual energy (E_j) will generate a lower communication cost (C_j). The node j with lowest value of C_j is the best choice for next hop.

Lines 13–26 of Algorithm 1 show that a new record for the destination is added in neighbor table if the distance from the neighbor node j to the destination ($D_{(j,Dst)}$) is less than the distance from the node i to the destination; that is, $D_{(j,Dst)}(hp) < D_{(i,Dst)}$.

A new record with the information of the neighbor node j is also added with the new calculated values as shown in Algorithm 1 from lines 27–38.

The neighbor table constructor algorithm repeats the same process of updating the neighbor table after receiving every new Hello packet.

5.3. Routing Services Module. The routing services module contains four submodules: QoS classifier, routing table constructor algorithm, routing table, and path selector algorithm. The QoS classifier submodule is responsible for categorizing the data packets into delay-sensitive packets (DSPs), reliability-sensitive packets (RSPs), and ordinary packets (OPs). The routing table constructor algorithm is used to construct and update the routing table. The routing table submodule stores the required information of the next hop(s) for the data packets. The routing table structure for node i is shown in Figure 5. The path selector algorithm chooses the best path(s) for each category (DSP, RSP, or OP) of traffic, based on the QoS requirement.

5.3.1. Routing Table Constructor Algorithm. The neighbor table entries are used to construct the routing table. Neighbor table contains multiple records for each destination. The routing table constructor algorithm determines the best next hops OPs, RSPs, and DSPs. It filters the neighbor table and only chooses an entry with the best values for the routing table. As shown in Algorithm 2, a new record is added in the routing table for each destination Dst $\in$ {MDC, NSC, BAN}. Lines 2–8, 9–27, and 28–34 are used to determine the values related to the OPs, RSPs, and DSPs, respectively.

The next hop for OPs (NH_E) will be the destination ID (ID_{Dst}) if the neighbor node is also the destination node (line 2). Otherwise a neighbor node j with the lowest communication cost (C_j) will be selected as next hop (NH_E).

For RSPs, the routing table constructor algorithm of ZEQoS finds three possible paths to ensure the minimum required reliability. For each destination, the three paths with highest reliabilities ($R_{path1(i,Dst)}$, $R_{path2(i,Dst)}$, and $R_{path3(i,Dst)}$) are chosen and their corresponding next hops (NH_{R1}, NH_{R2},

INPUT: Neighbor table, i's neighbor table records $\mathrm{NH}_{(i,\mathrm{Dst})}$, $\forall$ Dst $\in$ {MDC, NSC, BAN}

(1) **for** each destination Dst $\in$ {NSC, MDC, BAN} **do**
(2) **if** ($\mathrm{ID}_j(\mathrm{nt}) == \mathrm{ID}_{\mathrm{Dst}}(\mathrm{nt})$) **then**
(3) $\mathrm{NH}_E \leftarrow \mathrm{ID}_{\mathrm{Dst}}(\mathrm{nt})$
(4) **else**
(5) **if** ($C_j == \min_{k \in \mathrm{NH}_{(i,\mathrm{Dst})}} C_k$) **then**
(6) $\mathrm{NH}_E \leftarrow \mathrm{ID}_j(\mathrm{nt})$
(7) **end if**
(8) **end if**
(9) NH_R = {All neighbor nodes $j \in \mathrm{NH}_{(i,\mathrm{Dst})}$}
(10) **if** ($\mathrm{NH}_R == \mathrm{NULL}$) **then**
(11) Put NULL in $\mathrm{NH}_{R1}, \mathrm{NH}_{R2}, \mathrm{NH}_{R3}, R_{\mathrm{option1}(i,\mathrm{Dst})}, R_{\mathrm{option2}(i,\mathrm{Dst})}, R_{\mathrm{option3}(i,\mathrm{Dst})}$
(12) **else**
(13) Sort NH_R in descending order of $R_{\mathrm{path}(i,\mathrm{Dst})}$
(14) NH_{R1} = first neighbor node $j \in \mathrm{NH}_R$
(15) $R_{\mathrm{option1}(i,\mathrm{Dst})} = R_{\mathrm{path}(i,\mathrm{Dst})}$
(16) $P_{\mathrm{error}} = 1 - R_{\mathrm{option1}(i,\mathrm{Dst})}$
(17) **if** ($|\mathrm{NH}_R| > 1$)
(18) NH_{R2} = second neighbor node $j \in \mathrm{NH}_R$
(19) $P_{\mathrm{error}} = P_{\mathrm{error}} * (1 - R_{\mathrm{path}(i,\mathrm{Dst})})$
(20) $R_{\mathrm{option2}(i,\mathrm{Dst})} = 1 - P_{\mathrm{error}}$
(21) **end if**
(22) **if** ($|\mathrm{NH}_R| > 2$)
(23) NH_{R3} = third neighbor node $j \in \mathrm{NH}_R$
(24) $P_{\mathrm{error}} = P_{\mathrm{error}} * (1 - R_{\mathrm{path}(i,\mathrm{Dst})})$
(25) $R_{\mathrm{option3}(i,\mathrm{Dst})} = 1 - P_{\mathrm{error}}$
(26) **end if**
(27) **end if**
(28) NH = {All neighbor nodes $j \in \mathrm{NH}_{(i,\mathrm{Dst})}$}
(29) **if** ($|\mathrm{NH}| == 1$) **then**
(30) $\mathrm{NH}_D \leftarrow \mathrm{NH}$
(31) **else if** ($|\mathrm{NH}| > 1$) **then**
(32) Sort NH in ascending order of $\mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})}$
(33) NH_D = first neighbor node $j \in \mathrm{NH}$
(34) **end if**
(35) (add a new record for the Dst's information in the routing table)
(36) $\mathrm{ID}_{\mathrm{Dst}} \leftarrow \mathrm{ID}_{\mathrm{Dst}}(\mathrm{nt})$
(37) $L_{\mathrm{Dst}} \leftarrow L_{\mathrm{Dst}}(\mathrm{nt})$
(38) $\mathrm{NH}_E \leftarrow \mathrm{NH}_E$
(39) $\mathrm{NH}_{R1} \leftarrow \mathrm{NH}_{R1}$
(40) $\mathrm{NH}_{R2} \leftarrow \mathrm{NH}_{R2}$
(41) $\mathrm{NH}_{R3} \leftarrow \mathrm{NH}_{R3}$
(42) $\mathrm{NH}_D \leftarrow \mathrm{NH}_D$
(43) $\mathrm{NH}_{\mathrm{option1}(i,\mathrm{Dst})} \leftarrow \mathrm{NH}_{\mathrm{option1}(i,\mathrm{Dst})}$
(44) $\mathrm{NH}_{\mathrm{option2}(i,\mathrm{Dst})} \leftarrow \mathrm{NH}_{\mathrm{option2}(i,\mathrm{Dst})}$
(45) $\mathrm{NH}_{\mathrm{option3}(i,\mathrm{Dst})} \leftarrow \mathrm{NH}_{\mathrm{option3}(i,\mathrm{Dst})}$
(46) $\mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})} \leftarrow \mathrm{DL}_{\mathrm{path}(i,\mathrm{Dst})}$
(47) **end for**

Algorithm 2: Routing table constructor algorithm for ZEQoS.

and NH_{R3}) are stored in the routing table. The routing table constructor calculates and stores the three options for RSP. Line 9 of Algorithm 2 shows that the node i identifies the next hop candidates by searching the records which have the same $\mathrm{ID}_{\mathrm{Dst}}$ in neighbor table and stores them in the variable NH_R. If NH_R is empty, it means there is no next hop stored in NH_R. The node stores NULL to NH_{R1}, NH_{R2}, NH_{R3}, $R_{\mathrm{option1}(i,\mathrm{Dst})}$, $R_{\mathrm{option2}(i,\mathrm{Dst})}$, and $R_{\mathrm{option3}(i,\mathrm{Dst})}$. If NH_R is not empty, the next hop nodes' information is stored in the routing table one after another in descending order of their path reliabilities $R_{\mathrm{path}(i,\mathrm{Dst})}$. The first neighbor node j with the highest reliability in the routing table is stored as NH_{R1} (line 14). If there are two entries in NH_R then the aggregate reliability of first and second paths ($R_{\mathrm{option2}(i,\mathrm{Dst})}$) is calculated (lines 17–21). In case of more than two entries in NH_R, the aggregate reliability of first, second, and third

```
INPUT: Routing table, i's routing table records NH_(i,Dst), ∀ Dst ∈ {MDC, NSC, BAN}
(1)  for each data packet do
(2)   if data packet is delay-sensitive packet (DSP)
(3)     if (DL_path(i,Dst) <= DL_req) then
(4)       send to NH_D
(5)     else
(6)        drop the packet immediately
(7)     end if
(8)     else if data packet is reliability-sensitive packet (RSP)
(9)       if (R_option1(i,Dst) > R_req)
(10)         send to NH_R1
(11)       else if (R_option2(i,Dst) > R_req)
(12)         send to NH_R1 and NH_R2
(13)         else if (R_option3(i,Dst) > R_req)
(14)           send to NH_R1, NH_R2 and NH_R3
(15)           else
(16)            drop the packet immediately
(17)       end if
(18)     else if data packet is Ordinary Packet (OP)
(19)        send to NH_E
(20)         else
(21)        drop the packet immediately
(22)  end if
(23) end for
```

Algorithm 3: Path selector algorithm for ZEQoS.

paths ($R_{\text{option3}(i,\text{Dst})}$) is calculated (lines 22–26). In the routing table, the three paths with highest reliabilities ($R_{\text{path1}(i,\text{Dst})}$, $R_{\text{path2}(i,\text{Dst})}$, and $R_{\text{path3}(i,\text{Dst})}$) are chosen and their corresponding next hops (NH_{R1}, NH_{R2}, and NH_{R3}) are stored for each destination in the routing table. The routing table constructor calculates and stores the three options for RSP. The detailed calculations of $R_{\text{option1}(i,\text{Dst})}$, $R_{\text{option2}(i,\text{Dst})}$, and $R_{\text{option3}(i,\text{Dst})}$ are discussed in earlier work [6].

For DSP data, the path delay $\text{DL}_{\text{path}(i,\text{Dst})}$ has been calculated by using the neighbor table constructor algorithm (line 9 of Algorithm 1) and stored in neighbor table for each next hop candidate. The node stores the neighbor node's IDs in the variable NH (line 28). If NH has only one entry, this means there is only one path available. The node stores this entry to NH_D (line 30). Otherwise the node sorts the NH entries in ascending order with respect to the path delay (i.e., $\text{DL}_{\text{path}(i,\text{Dst})}$) values and then stores the first entry which has the lowest path delay in NH_D (lines 32-33). The next hop candidate NH_D is then stored with its path delay value ($\text{DL}_{\text{path}(i,\text{Dst})}$) in the routing table. Algorithm 2 (lines 27–38) shows that a new record for the destination Dst is added with the calculated values.

The routing table constructor algorithm repeats the same process of updating the routing table after receiving every new Hello packet.

5.3.2. Path Selector Algorithm. The data packets from both upper layers and packet classifiers are received by QoS classifier. The QoS classifier classifies the packets into DSP, RSP, and OP data. For each data packet, the path selector algorithm checks the QoS requirement and chooses the most appropriate next hop(s). Lines 2–7, 8–17, and 18–21 of Algorithm 3 are used for the selection of appropriate next hops of DSPs, RSPs, and OPs, respectively. The path selector algorithm compares the delay requirement (DL_{req}) with the path delay ($\text{DL}_{\text{path}(i,\text{Dst})}$) of NH_D which is stored in the routing table. If the path delay ($\text{DL}_{\text{path}(i,\text{Dst})}$) is lower than required delay (DL_{req}), the packet is sent to NH_D (lines 3-4). Otherwise, the packet is dropped (line 6).

For RSPs, the path selector algorithm checks if the reliability of a single path exceeds R_{req}; then a single path is used to send these packets through NH_{R1} (lines 9-10). In case the required reliability is greater than the reliability of any single path, then, the path selector selects two paths (by using NH_{R1} and NH_{R2}) whose aggregate reliability is more than the requested R_{req} (lines 11-12). If not, three paths are used as long as their aggregate reliability is greater than the R_{req} (lines 13-14) or else the packet is dropped. The method of finding the aggregate reliabilities is given in Algorithm 2 and discussed in Section 5.3.1. For OPs, the path selector algorithm returns the next hop NH_E (lines 18-19). Any unknown packet should be dropped without assigning any next hop (line 21).

5.4. QoS-Aware Queuing Module. The data packets are sent to the QoS-aware queuing module (QQM) after the selection of appropriate next hop(s) by routing services module. QQM receives the data packets and separates these packets in three classes (DSPs, RSPs, and OPs). An individual queue is used for each class of packets. QQM functions are the same as discussed in [17]. The priority of the DSPs queue is higher than that of the RSPs and OPs queues. The RSPs queue has lower priority than DSPs queue. The priority of OPs queue is

Table 3: Parameters information.

Deployment	Area	16 m by 21 m
	Deployment type	Movable source node BAN_2 (shown in Figure 6)
	Number of nodes	49 nodes (24 BANs, 24 MDCs, and 1 NSC)
	Initial nodes locations	As shown in Figure 6
	Initial node energy	18720 J (=2 AA batteries)
	Buffer size	32 packets
	Link layer trans. rate	250 Kbps
	Transmit power	−25 dBm
Task	Application type	Event-driven
	Max. packet size	32 Bytes
	Traffic type	CBR (constant bit rate)
MAC	IEEE 802.15.4	Default values
Simulation	Time	2003 seconds (3 seconds is setup time)

the lowest. By default, the DSPs queue with highest priority sends the packets first. The packets from lower priority RSP queue will be sent only when the DSPs queue is empty. The OPs need to wait until the DSPs and RSPs queues are empty. However, for fair treatment of OPs data, a timeout is used by all the queues. A queue sends the packets to the MAC layer within the period specified by the timeout for that queue. QQM changes the control from higher priority queue to lower priority queue after the queue timeout occurs.

6. Performance Evaluation

OMNeT++ based simulator Castalia [26] is used to test the performance of the proposed ZEQoS routing protocol. The simulation results prove that the ZEQoS approach based on end-to-end path delays and reliabilities in addition to the available energy and geographic information of the node are more effective for all data types (i.e., OPs, DSPs, and RSPs) when compared with DMQoS and noRouting protocols. The simulations are done by considering a real 24-bed hospital scenario outlined in Section 6.1. The details about the scenario, parameters information, and performance results for the simulations are provided below.

6.1. 49 Nodes in Hospital Environment. A real 24-patient bed hospital with a movable source node is considered for the testing of ZEQoS routing protocol, as shown in Figure 6. The approximate measurements used for this hospital environment are similar to the hematology-oncology unit of the children's hospital named IWK Health Centre Halifax, NS, Canada. The approximate area covered by this unit is 16 m by 21 m. The distance between two beds is 3 meters which is a recommended transmission range for BAN communication in hospital environment. Each BAN transmits the data to its respective MDC. All the BANs and MDCs are sending or receiving Hello protocols to/from other nodes and the NSC. The total numbers of nodes used in this scenario are 49 which include 24 BANs, 24 MDCs, and 1 NSC. The NSC is placed on the left side of the deployment area. The patient rooms are in four rows. The room numbers 1–7, 8–12, 13–17, and 10–24 are in rows 1, 2, 3, and 4, respectively. Room number 18 and the nursing station are just in front of all these rows.

The MDCs and BANs are movable but normally an MDC placed in a room moves only within that room. BANs can move freely anywhere. It is assumed that the MDC of one room has a connection with the MDC of the next room. The patient node BAN_2 is considered as a movable BAN coordinator (BANC). As a fast walking patient, the speed of movable BANC is set to 1 meter per second. The node BAN_2 moves vertically as shown by the green arrows in Figure 6. The source node BAN_2 displays its data to MDC_2.

6.2. Parameters Used for Simulations. The transmit power used in simulations is −25 dBm. The transmission range of −25 dBm is about 3 meters which is the recommended value for BAN communication [2] in hospital environment. The network parameters used in our simulations are shown in Table 3.

6.3. Performance Results. The source nodes send a total of 95 K data packets in the 49-node hospital environment. The above mentioned parameters are calculated after the transmission of every 9.5 K packet of all types sent by the source nodes. All types of data packets including OPs, DSPs, and RSPs are sent from source nodes. To achieve a 97% confidence interval for the illustrative results, three runs are simulated in every experiment which may introduce a maximum error of 3×10^{-3}, based on the error calculation done by Castalia simulator [27]. The below two cases are considered for the same scenario shown in Figure 6.

Case 1. A fixed number of DSPs and RSPs but a variable number of OPs are sent from the source nodes. The number of DSPs is 1.2 K when 9.5 K packets are sent by source nodes. After that 7 K DSPs are consistently included in the offered traffic loads by source nodes. The 7 K RSPs are included consistently in all offered traffic loads. The OPs are continuously increased from 1 K to 81 K as with the increase of offered traffic load from 9.5 K to 95 K, respectively. The types of data packets included in the offered traffic load are shown in Figure 7(a).

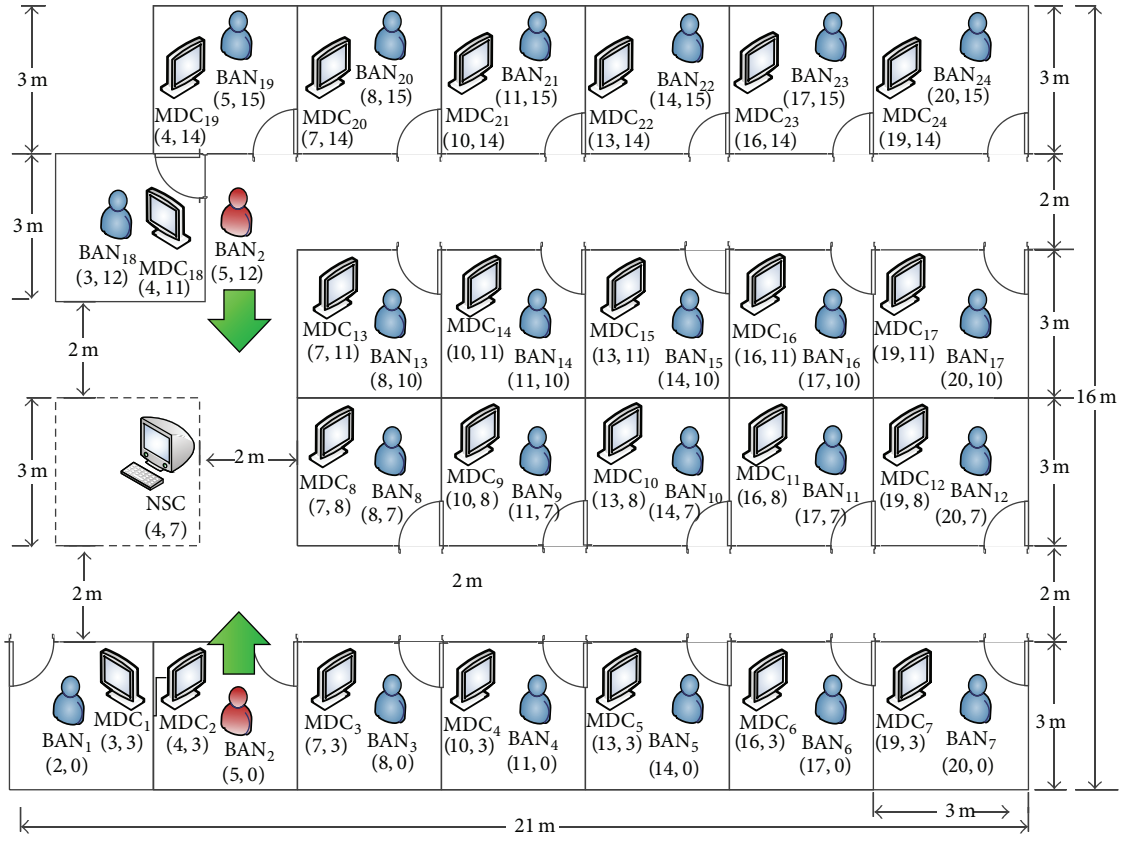


FIGURE 6: Node deployment for 24 patient beds in hospital environment.

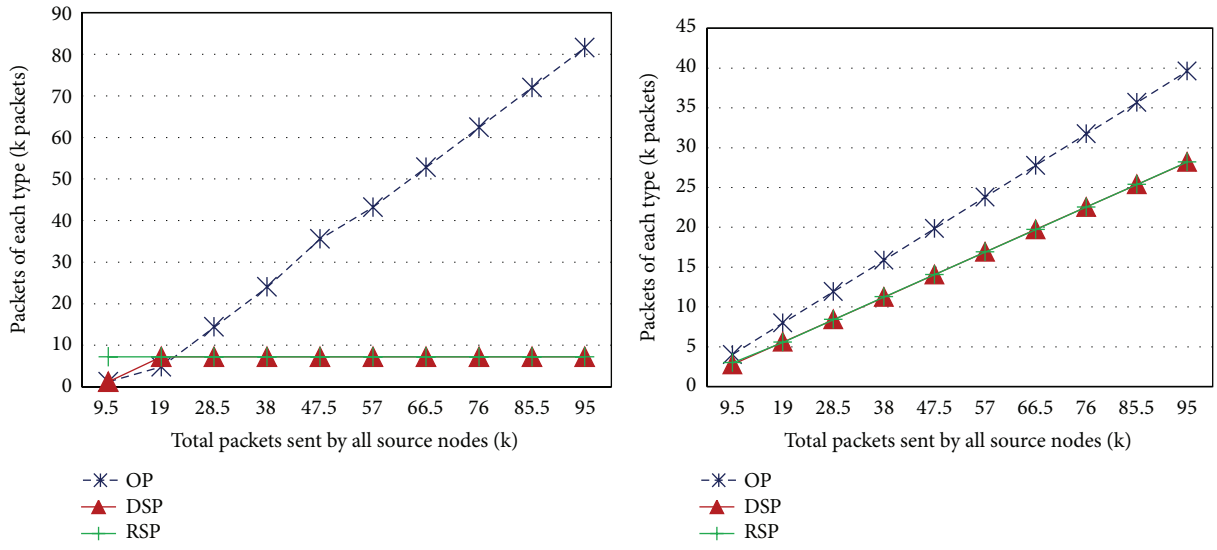
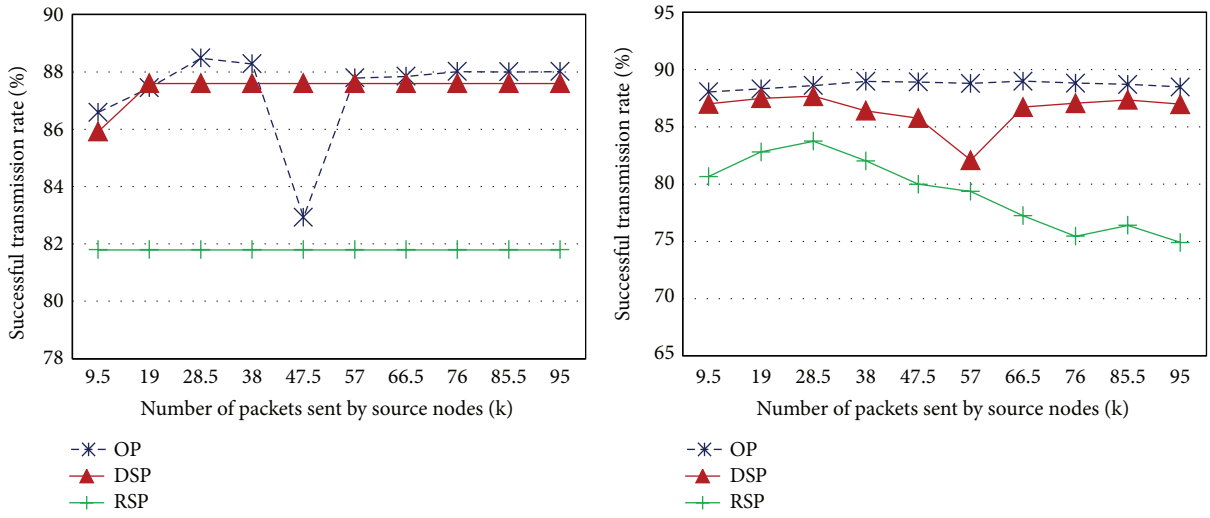


FIGURE 7: Offered traffic by source nodes.

Case 2. A variable number of OPs, DSPs, and RSPs are sent with the ratio of 40%, 30%, and 30%, respectively. The OPs constitute from 4 K to 39.5 K packets as the offered traffic load is increased from 9.5 K to 95 K. Similarly, DSPs and RSPs packets constitute from 2.8 K to 28 K packets of each type, when the total offered traffic load by source nodes is increased from 9.5 K to 95 K packets. Figure 7(b) shows the types of packets included in the offered traffic load for Case 2.

The throughput, packets forwarded by intermediate nodes, network traffic, packets dropped at the network layer,



(a) Case 1 (b) Case 2

FIGURE 8: Throughput versus offered traffic.

packets dropped on MAC layer, and energy consumption are measured. The performance results of each parameter are discussed below.

6.3.1. Throughput. The throughput is measured by calculating the number of packets received successfully at the destination nodes. The successful transmission rate or throughput is measured after the transmission of every 9.5 K packet sent by the source. For Case 1, Figure 8(a) shows that ZEQoS provides a consistent reliability which is in excess of 82%, 85%, and 81% for OPs, DSPs, and RSPs, respectively. For Case 2, as shown in Figure 8(b), the successful transmission rate of OPs, DSPs, and RSPs is in excess of 88%, 86%, and 75%, respectively.

The results from Figure 8 show that the mechanism of ZEQoS handles all the data types (i.e., OPs, DSPs, and RSPs) successfully with higher throughput. ZEQoS overcomes the issues of traffic congestion by using the end-to-end path delays and reliabilities for DSPs and RSPs, respectively. Also the transmission of RSPs over redundant paths ensures the higher reliability of RSPs packets.

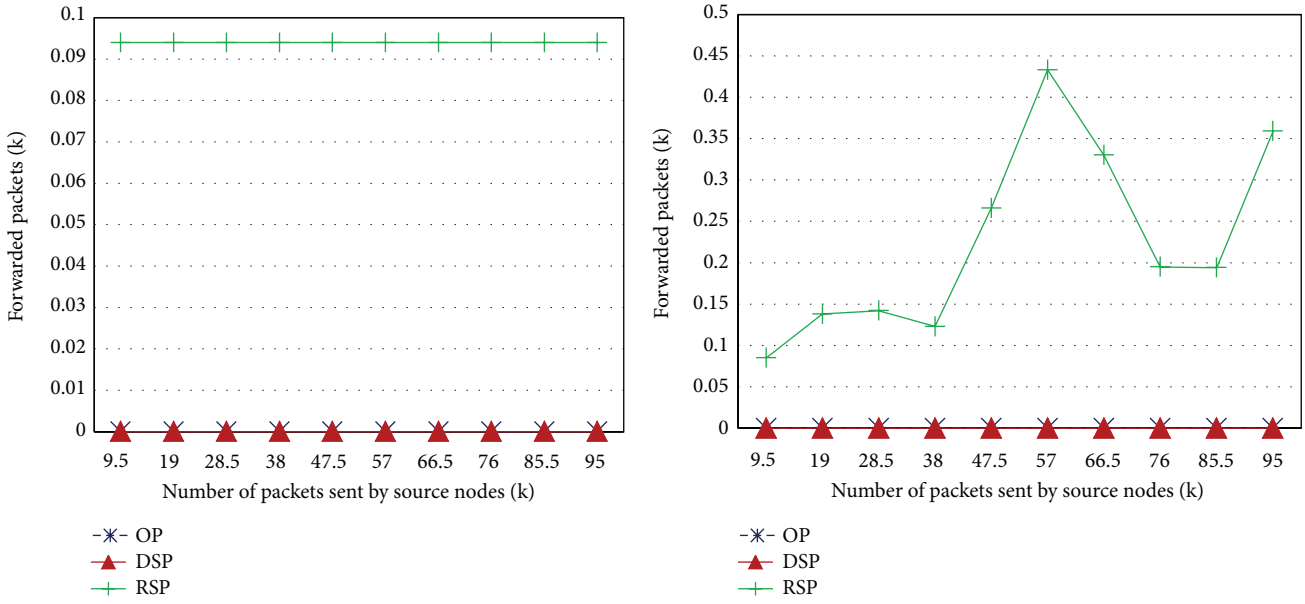
The path selection mechanism of ZEQoS considers the geographic location, energy availability, end-to-end path delays, and end-to-end path reliabilities for all nodes in the network which helps improve the overall throughput for all the data types.

6.3.2. Packets Forwarded by Intermediate Nodes. The approach used in ZEQoS for the selection of the most appropriate next hop is very effective. In the proposed ZEQoS scheme, a BAN coordinator does not send data to other BAN coordinators unless it is necessary. The BAN coordinator in the proposed ZEQoS sends data to another BAN coordinator only if it is necessary. The BAN coordinators send the data packets directly to the destinations. In noRouting, the delay-sensitive data packets are forwarded to random next hop devices instead of algorithm's next hop based on end-to-end path delay routes. Figure 9 shows the number of OPs, DSPs, and RSPs forwarded by the intermediate nodes.

It is seen from Figure 9 that no OPs or DSPs data packets are forwarded by any intermediate nodes. In Case 1, the number of RSPs forwarded by intermediate nodes is only 94 which is negligible when compared to the overall network traffic. In Case 2, from Figure 9(b) it is shown that the intermediate nodes forwarded 85 to 433 RSPs when offered traffic is increased from 9.5 K to 95 K. The control of Hello packets broadcast also helps reduce the packets forwarded by intermediate nodes.

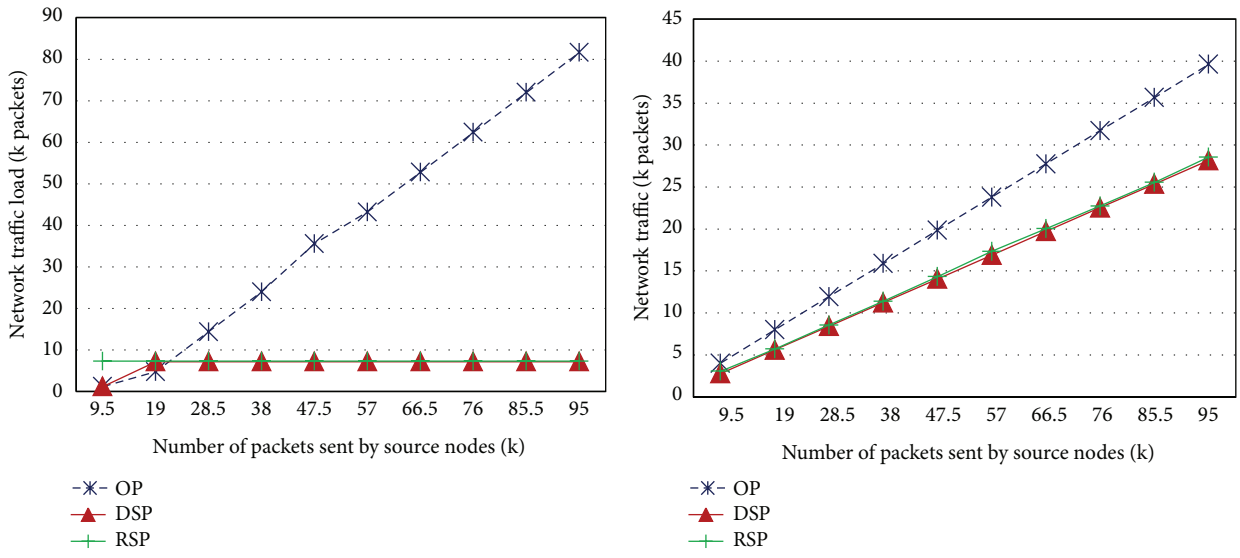
6.3.3. Overall Network Traffic. The lower number of forwarded packets as discussed in the previous section helps reduce the overall network traffic. The Hello packets are not added in this network traffic. In Case 1, Figure 10(a) shows that the overall network traffic due to OPs, DSPs, and RSPs are almost 7 K, 7 K, and 1 K to 81 K, respectively. The numbers of Hello packets are 179 K to 2198 K when 9.5 K to 95 K offered traffic load is applied from the source nodes, respectively. In Case 2, the overall network traffic due to OPs, DSPs, and RSPs is almost 4 K to 39 K, 2.5 K to 28 K, and 3 K to 28.5 K, respectively, as shown in Figure 10(b). In addition to data packets, 182 K to 2171.5 K Hello packets are also part of overall network traffic when 9.5 K to 95.5 K packets are sent by source nodes, respectively.

6.3.4. Packets Dropped at the Network Layer. In previous protocols like DMQoS [17], the source nodes calculate the hop-by-hop delay and reliability of the next hop nodes for the DSPs and RSPs, respectively, and send the data to the best next hop which has lowest delay for DSPs and highest reliability for RSPs. The next hop then calculates the delays or reliabilities of its upstream nodes. The packets are dropped in case of not meeting the requested delay or reliability by all neighboring upstream nodes. ZEQoS resolves this problem by using the



(a) Case 1 (b) Case 2

Figure 9: Packets forwarded by intermediate nodes.

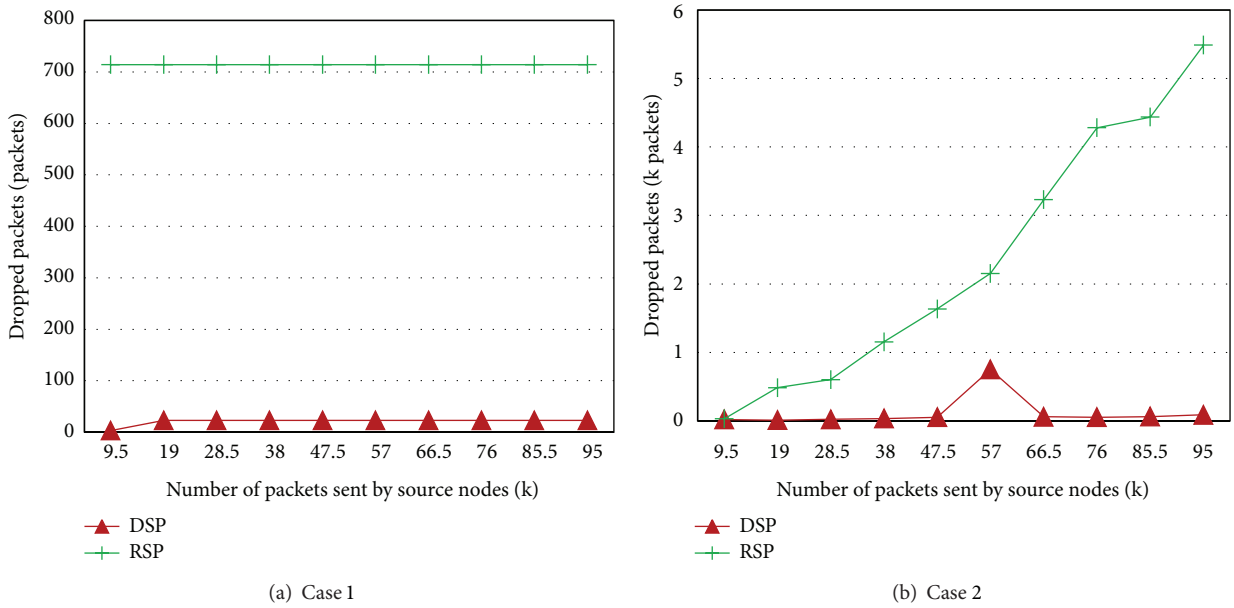


(a) Case 1 (b) Case 2

Figure 10: Overall network traffic versus offered load.

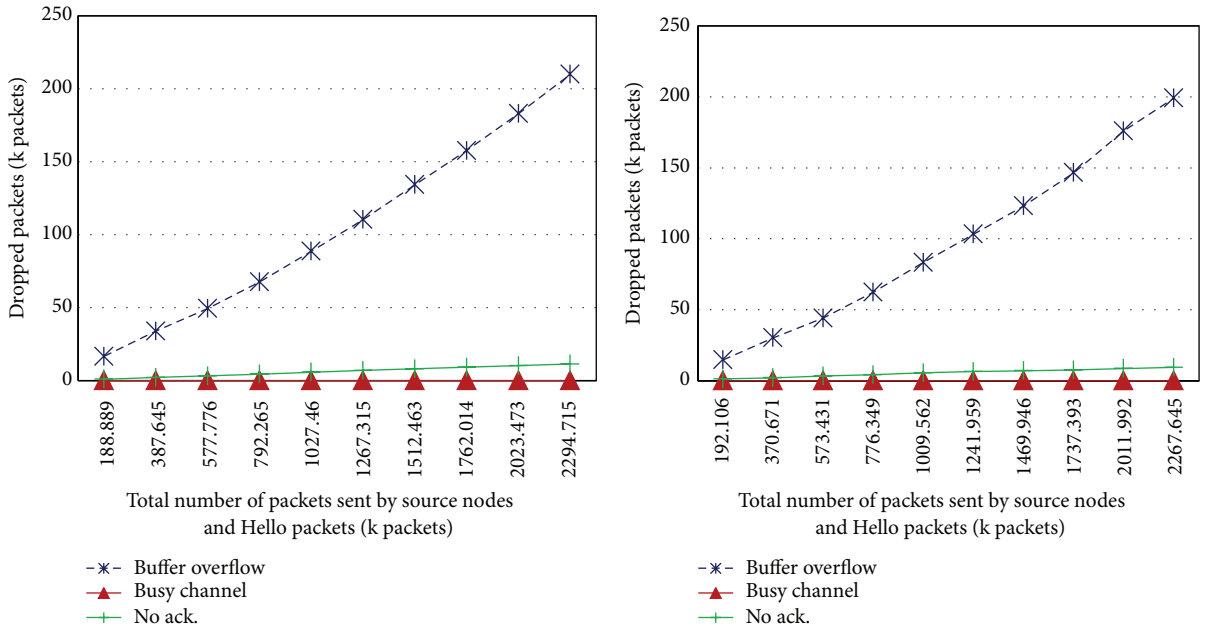
end-to-end path delays and reliabilities for DSPs and RSPs, respectively. Also the use of three redundant paths for RSPs in ZEQoS ensures better transmission rate. In Case 1, ZEQoS drops 23 DSPs and 714 RSPs data packets for all the traffic loads as shown in Figure 11(a). In Case 2, Figure 11(b) shows that the DSPs and RSPs dropped at the network layer due to not meeting the requested reliability and delay requirements are an average of 0.2% and 4.4%, respectively.

6.3.5. Packets Dropped by the MAC Layer. The total number of packets dropped by the MAC layer due to buffer overflow, busy channel, and no acknowledgements is measured. Figure 12 shows the packets dropped by MAC layer for Cases 1 and 2, respectively. The total offered traffic including Hello packets is 188 K to 2294 K and 192 K to 2267 K for Cases 1 and 2, respectively. No data packets are dropped due to busy channel in both cases. Also the packets dropped due to no acknowledgments increases from 1 K to 11 K in both cases. It is seen from Figure 12 that packets dropped due to the MAC buffer overflow are very high. In Case 1, the packets dropped due to the buffer overflow are 16 K to 209 K, whereas, in Case 2, the average packets dropped due to buffer overflow is 8.4%.



(a) Case 1 (b) Case 2

Figure 11: Packets dropped at the network layer due to lower delay or reliability requirements.



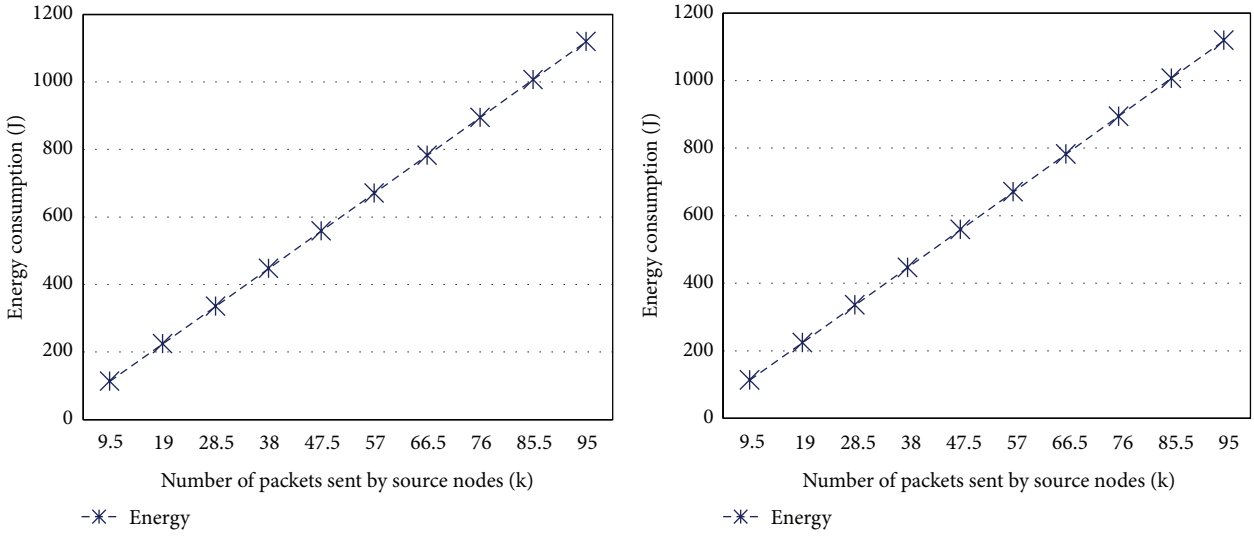
(a) Case 1 (b) Case 2

Figure 12: Packets dropped by the MAC layers.

6.3.6. Overall Energy Consumption. The overall energy consumption in both cases for ZEQoS is discussed in this section. It shows that ZEQoS provides a consistent and more reliable delivery of all three types of data packets (OPs, DSPs, and RSPs) as previously discussed in Section 6.3.1. The energy consumptions of both cases are similar as shown in Figure 13. The figure shows that ZEQoS consumes 112 to 118 Joules of energy when the offered load is 9.5 K to 95 K data packets as sent by source nodes. The drawback of ZEQoS is to consume much higher energy as compared to the energy consumption of the protocols (EPR, QPRD, and QPRR) which are not handling all three data types OPs, DSPs, and RSPs at a time.

7. Performance Comparison with DMQoS and NoRouting

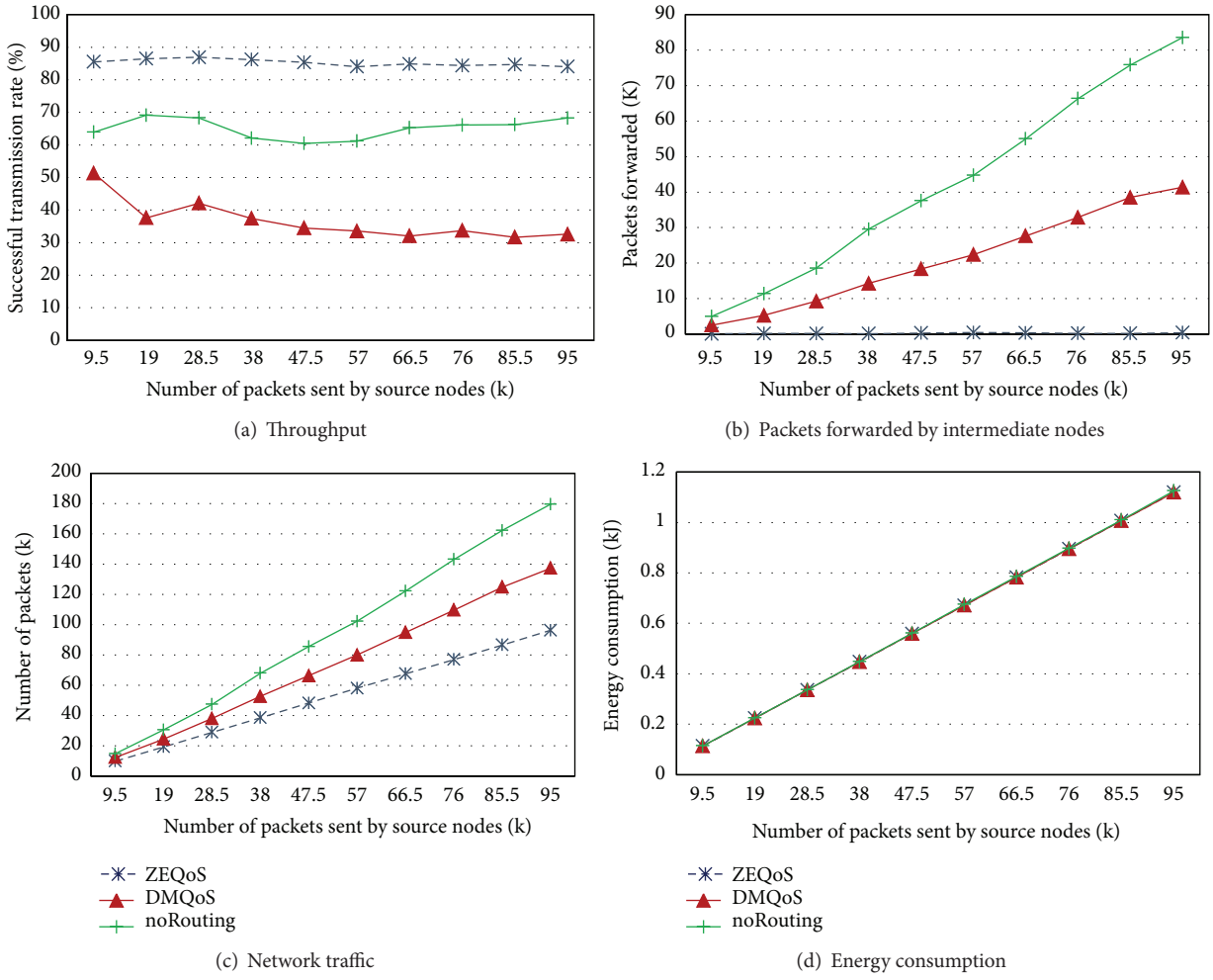
In this section, the performance of ZEQoS is compared with the DMQoS routing protocol [17] and noRouting. Norouting mechanism is used in the noRouting is case. The packets are



(a) Case 1

(b) Case 2

Figure 13: Overall energy consumption.



(a) Throughput

(b) Packets forwarded by intermediate nodes

(c) Network traffic

(d) Energy consumption

Figure 14: Performance comparison for different parameters.

forwarded to random next hop devices instead of following any algorithm's next hop. The comparison with noRouting is used to verify whether forwarding the packets to a random next hop device results in a better successful transmission rate than the ZEQoS routing which is based on energy and QoS-aware algorithm. The experimental results, shown in Figure 14, prove that the approach used by ZEQoS is more effective. The node deployment used for this test is similar to Case 2 of Section 6.3. The network parameters used in our simulations are shown in Table 2. The offered traffic load generated from nodes is 40%, 30%, and 30% of OPs, DSPs, and RSPs, respectively. The total 95 K packets are sent from the source nodes and the results are noted after every 9.5 K packet.

Figure 14(a) shows that ZEQoS provides a consistent 84% throughput; however, it is seen that the reliabilities of DMQoS and noRouting are on average 36% and 65%, respectively. Figure 14(b) shows that the packets forwarded by the intermediate nodes in ZEQoS are almost negligible; whereas, DMQoS and noRouting forward 21 K and 42 K packets, respectively. The hop-by-hop mechanism used in DMQoS causes the increased forwarded packets. The network traffic is increased when more packets are forwarded by intermediate nodes as shown in Figure 14(c). The increased network traffic causes the traffic congestion and more packets are dropped on MAC and network layers as explained in Sections 6.3.4 and 6.3.5. It is seen from Figure 14(d) that the energy consumption for all three protocols is the same for all the traffic loads.

8. Conclusion

A new modular energy and QoS-aware routing protocol (ZEQoS) for hospital BAN communication is proposed in this paper. The modules of new protocol are divided into two main types: MAC layer modules and network layer modules. MAC layer modules include the MAC receiver, the reliability module, the delay module, and the MAC transmitter. The packet classifier, the Hello protocol module, the routing services module, and the QoS-aware queuing module are included in network layer modules.

The proposed ZEQoS routing protocol provides a mechanism with the help of neighbor table constructor algorithm, routing table constructor algorithm, and path selector algorithm to calculate the communication costs, end-to-end path delays, and end-to-end path reliabilities of all possible paths from a source to destination and then decides on the best possible path(s) with the consideration of QoS requirement of the OPs, RSPs, and DSPs.

OMNeT++ based simulator Castalia 3.2 [26] was used to test the performance of the proposed protocol. The simulations were performed by considering a real hospital scenario when a source node was movable. All three types of data packets OPs, RSPs, and DSPs were sent from the source nodes. Both fixed and variable numbers of OPs, DSPs, and RSPs were considered. The simulation results showed that the ZEQoS had in excess of 81% and 75% throughput for all classes of packets in fixed and variable cases, respectively, when offered traffic load of 9.5 K to 95 K packets was used. The simulation results showed that the ZEQoS had superior performance in excess of 84% throughput when compared with DMQoS and noRouting provides 36% and 65%, respectively.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

References

[1] P. J. Riu and K. R. Foster, "Heating of tissue by near-field exposure to a dipole: a model analysis," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 8, pp. 911–917, 1999.

[2] "IEEE 802.15 WPAN Task Group 6 (TG6) Body Area Networks," IEEE 802.15, November 2007, http://www.ieee802.org/15/pub/TG6.html.

[3] S. Ullah, H. Higgins, B. Braem et al., "A comprehensive survey of wireless body area networks on PHY, MAC, and network layers solutions," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1065–1094, 2012.

[4] Z. A. Khan, *Advanced Zonal Rectangular LEACH (AZR-LEACH): an energy efficient routing protocol for wireless sensor networks [Ph.D. thesis]*, Dalhousie University, Halifax, Canada, 2012.

[5] Z. Khan, S. Sivakumar, W. Phillips, and B. Robertson, "QPRD: QoS-aware peering routing protocol for delay sensitive data in hospital body area network communication," in *Proceedings of the 7th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA '12)*, pp. 178–185, Victoria, Canada, November 2012.

[6] Z. Khan, S. Sivakumar, W. Phillips, and B. Robertson, "A QoS-aware routing protocol for reliability sensitive data in hospital body area networks," *Procedia Computer Science*, vol. 19, pp. 171–179, 2013.

[7] T. Lu and J. Zhu, "Genetic algorithm for energy-efficient QoS multicast routing," *IEEE Communications Letters*, vol. 17, no. 1, pp. 31–34, 2013.

[8] D. Djenouri and I. Balasingham, "Traffic-differentiation-based modular QoS localized routing for wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 6, pp. 797–809, 2011.

[9] X. Liang and I. Balasingham, "A QoS-aware routing service framework for biomedical sensor networks," in *Proceedings of the 4th IEEE International Symposium on Wireless Communication Systems (ISWCS '07)*, pp. 342–345, Trondheim, Norway, October 2007.

[10] K. Zeng, K. Ren, W. Lou, and P. J. Moran, "Energy aware efficient geographic routing in lossy wireless sensor networks with environmental energy supply," *Wireless Networks*, vol. 15, no. 1, pp. 39–51, 2009.

[11] X. Liang, I. Balasingham, and S.-S. Byun, "A reinforcement learning based routing protocol with QoS support for biomedical sensor networks," in *Proceedings of the 1st International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL '08)*, Aalborg, Denmark, October 2008.

[12] M. A. Razzaque, M. M. Alam, M. Mamun-Or-Rashid, and C. S. Hong, "Multi-constrained QoS geographic routing for heterogeneous traffic in sensor networks," *IEICE Transactions on Communications*, vol. E91-B, no. 8, pp. 2589–2601, 2008.

[13] D. Djenouri and I. Balasingham, "New QoS and geographical routing in wireless biomedical sensor networks," in *Proceedings of the 6th International Conference on Broadband Communications, Networks and Systems (BROADNETS '09)*, Madrid, Spain, September 2009.

[14] S. Wu and K. S. Candan, "Power-aware single- and multipath geographic routing in sensor networks," *Ad Hoc Networks*, vol. 5, no. 7, pp. 974–997, 2007.

[15] E. Felemban, C.-G. Lee, and E. Ekici, "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. 738–753, 2006.

[16] T. He, J. A. Stankovic, C. Lu, and T. F. Abdelzaher, "A spatiotemporal communication protocol for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 10, pp. 995–1006, 2005.

[17] A. Razzaque, C. S. Hong, and S. Lee, "Data-centric multiobjective QoS-aware routing protocol for body sensor networks," *Sensors*, vol. 11, no. 1, pp. 917–937, 2011.

[18] M. Chen, T. Kwon, and Y. Choi, "Energy-efficient differentiated directed diffusion (EDDD) in wireless sensor networks," *Computer Communications*, vol. 29, no. 2, pp. 231–245, 2006.

[19] M. Chen, V. C. M. Leung, S. Mao, and Y. Yuan, "Directional geographical routing for real-time video communications in wireless sensor networks," *Computer Communications*, vol. 30, no. 17, pp. 3368–3383, 2007.

[20] X. Huang and Y. Fang, "Multiconstrained QoS multipath routing in wireless sensor networks," *Wireless Networks*, vol. 14, no. 4, pp. 465–478, 2008.

[21] M. Chen, T. Kwon, S. Mao, Y. Yuan, and V. Leung, "Reliable and energy-efficient routing protocol in dense wireless sensor networks," *International Journal on Sensor Networks*, vol. 4, no. 1-2, pp. 104–117, 2008.

[22] M. Chen, V. C. M. Leung, S. Mao, Y. Xiao, and I. Chlamtac, "Hybrid geographic routing for flexible energydelay tradeoff," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 4976–4988, 2009.

[23] Z. Khan, S. Sivakumar, W. Phillips, and N. Aslam, "A new patient monitoring framework and Energy-aware Peering Routing Protocol (EPR) for Body Area Network communication," *Journal of Ambient Intelligence and Humanized Computing*, 2013.

[24] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: a survey," *Mobile Networks and Applications*, vol. 16, no. 2, pp. 171–193, 2011.

[25] J. Xu, W. Liu, F. Lang, Y. Zhang, and C. Wang, "Distance measurement model based on RSSI in WSN," *Wireless Sensor Network*, no. 2, pp. 606–611, 2010.

[26] NICTA, "Castalia," National ICT Australia, March 2011, https://castalia.forge.nicta.com.au/index.php/en/.

[27] NICTA, "Castalia, Wireless Sensor Network Simulator," May 2014, http://castalia.research.nicta.com.au/index.php/en/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 709135, 10 pages
http://dx.doi.org/10.1155/2014/709135

Research Article

Energy Consumption Optimisation for Duty-Cycled Schemes in Shadowed Environments

Tatjana Predojev,[1] Jesus Alonso-Zarate,[1] Mischa Dohler,[2] and Luis Alonso[3]

[1] *Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Parc Mediterrani de la Tecnologia, Building B4, Avenida Carl Friedrich Gauss 7, Castelldefels, 08860 Barcelona, Spain*
[2] *Department of Informatics, King's College London, Strand, London WC2R 2LS, UK*
[3] *Universitat Politècnica de Catalunya, Parc Mediterrani de la Tecnologia (PMT), Esteve Terrades 1, Castelldefels, 08860 Barcelona, Spain*

Correspondence should be addressed to Tatjana Predojev; tatjana.predojev@cttc.es

Received 31 December 2013; Accepted 14 May 2014; Published 28 May 2014

Academic Editor: Sana Ullah

The focus of this study is the optimal configuration of a wireless low-power duty-cycled network with respect to the minimal energy consumption. Precisely, the energy consumption of a truncated-ARQ scheme in realistic shadowing environments is examined for the reference IEEE 802.15.4e standard protocol and for its cooperative extension that is presented in the paper. We show how to choose between the direct or multihop forwarding and the cooperative version of the two. We determine the optimal forwarding strategy for both loose and strict reliability requirements. Low-power links are parametrised by the interdevice distance and the corresponding outage probability, for the fixed output transmission power. It is shown that significant amounts of energy can be saved when the most adequate scheme of the three is applied. All analytical results are validated in the network simulator ns-3.

1. Introduction

A device domain of the machine-to-machine (M2M) communication system largely consists of resource-constrained, low-power, and energy-efficient devices. Following years of research and fine-tuning, viable technical solutions for achieving the adequate low energy regime have been devised and the standardised protocol stack has been put forward [1]. The lifetime of wireless M2M devices is measured in years or decades; extreme energy efficiency is thus a must and only achieved through aggressive duty-cycling [2]. A duty-cycled device keeps the radio transceiver in sleep state most of the time, except for the periodic wake-ups used to transmit the collected data. In this way, both overhearing and idle listening are evaded with the goal of conserving energy. The IEEE 802.15.4e standard amendments [3] define the required duty-cycled scheme. Arguments for applying this scheme are provided in [1] that examines the most energy-efficient solutions over the entire protocol stack. With the standardised protocol stack in place, remaining work is to find the adequate device configuration which includes the optimal forwarding strategy, in realistic environments. These issues are addressed in our study.

Link (un)reliability is core to our study: it is known that short interdevice distances typically imply more reliable links, whilst reliability decreases with the increase in distance. Therefore, one of the key parameters in our work is the optimal interdevice distance under the typical (low) values of output transmission power. We characterise the resulting link (un)reliability with the link outage probability. Link outages result in discarded packets; therefore, outage probability provides the estimation of link quality. In addition, this approach enables network design under outage constraints specified in advance.

In order to formulate the energy consumption model for the protocol stack of M2M low-power devices, we start by deducing a link metric that considers realistic operating conditions. Physical layer studies (e.g., [4]) typically focus on the physical phenomena of the wireless channel and related effects on the error probability. The channel is thus subject

to large-, medium-, and/or small-scale fading, with the latter two in time being static, block, or fast. On the other hand, the studies of upper layers experimentally measure the impact of wireless channel, such that the channel effects are reflected in bursty or independent link behaviour (e.g., [5]). We connect the two approaches into the link layer analytical model, while capitalising on the results and observations of previous works. Indeed, we analyse how wireless channel effects interact with higher layers, link layer in particular, and how this interaction affects the link quality. A metric we propose is dependent on two critical parameters under study: link distance and the related outage probability. This metric is the Average Number of Transmissions per Packet $\overline{N}_{tx}$. By considering outages at the link layer, we provide original approach in the analysis of low-power, unreliable links.

Accurate link characterisation offers insight into how to optimise the overall energy consumption. For example, dynamic forwarding is an effective way of combating link outages through path diversity. Using other available links when primary link is in outage eventually saves energy. Specifically, we focus on the Cooperative Automatic Repeat reQuest (C-ARQ) as a reactive form of dynamic forwarding. Traditionally, C-ARQ relies on overheard packets by the neighbouring devices which then become relays [6]. In a duty-cycled scheme, this is not possible. Therefore, we specifically adapt the C-ARQ technique to the duty-cycled scheme without assuming overhearing at the relay and optimise it for the most energy-efficient operating regime. The resulting scheme is denoted as Cooperative and Duty-Cycled ARQ (CDC-ARQ). The main idea of our CDC-ARQ scheme is to introduce path diversity in the scheduling functions. Opposed to C-ARQ, CDC-ARQ does not rely on overhearing, but rather on the analysis of wireless low-power links. We consider realistic wireless channel with shadow fading. In the analysis, we focus on the specifics of low-power M2M networks in order to present customised results that are ready applicable in practice. Nevertheless, our analytical model supports changes in the modulation or coding scheme, output transmission power, and so forth. One of the key features of CDC-ARQ technique is that it can be easily fitted into the standard, as no changes must be done to the physical (PHY) nor the medium access control (MAC) layers but just to the scheduling functions. All references to the standard in the paper refer to IEEE 802.15.4e [3].

Previously, we developed this idea in [7, 8]. In [7], the star topology is examined and it is shown that benefits can be obtained by forwarding through a relay after the initial transmission failure. In [8], we establish a cooperative communication scheme applicable to any topology and evaluate the scheme's energy consumption. In the present paper, we extend the analysis beyond cooperative scenario to offer a comprehensive overview of low-power wireless links. For a given outage probability constraint, we find the most energy-efficient forwarding strategy of the three available choices: direct, multihop, or CDC-ARQ forwarding. CDC-ARQ for a duty-cycled device alternates between the direct and multihop forwarding depending on the channel conditions.

In summary, the main contributions of this paper are as follows.

(1) A link energy consumption model is formulated to reflect the wireless channel effects.

(2) Link selection guidelines are provided which strive to minimise the overall energy consumption, for either loose or strict reliability requirements. In particular, the bounds for the efficient direct, multihop, or CDC-ARQ forwarding are derived and presented.

Finally, the analytical model provided in this paper is validated in ns-3 network simulator [9]. An ns-3 simulation mimics the real world as close as possible, since the implementation closely follows the related standard technology. Therefore, aside from validating our analytical model, we show that the techniques presented here are suitable for real devices and can be easily integrated into IEEE 802.15.4e standard.

The remainder of the paper is organised as follows: Section 2 lists some related works. Section 3 presents the system model. Section 4 contains the derived analytical energy model. The model validation is presented in Section 5 together with the results extended beyond the model in the simulations. Finally, the paper is concluded in Section 6.

2. Related Work

The optimal link distance that maximises energy efficiency has been previously investigated in [10] for different node densities and path loss exponents. The results obtained in [10] apply to a circular coverage area without considering the fading effects, which are acknowledged in this work. With the distance fixed, various cooperative schemes have been put forward in the literature in order to improve the reliability without trading it for higher energy consumption. Vardhe et al. study cooperation using distributed space time codes in [11], for equidistant relays on a direct path to destination. In [12], the energy efficiency of direct, multihop, and cooperative transmission schemes is studied for fixed outage probability in order to find the optimal output transmission power; the results, however, span the range of output power values significantly above the typical setting for the M2M low-power networks. These works apply to nonduty-cycled schemes and thus assume overhearing as the basis for cooperation which makes them unsuitable for duty-cycled systems envisioned in [3].

To overcome the complexity of cooperative scheme implementation at the PHY layer (such as synchronisation issues), cooperation at the link layer presents an alternative in the form of C-ARQ. In a C-ARQ scheme, a device seeks cooperation from neighbours to reroute data packets locally in the case of a temporary wireless channel outage on the primary link. C-ARQ for nonduty-cycled schemes was analytically studied in [13]. Alizai et al. take an experimental approach in [14] to show that a rerouting technique decreases the total number of packet (re)transmissions in low-power networks. A detailed energy consumption analysis is still needed to quantify the actual benefits and, therefore, configure

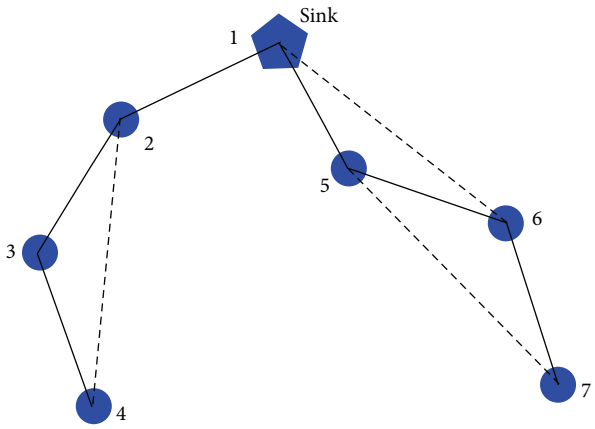


Figure 1: M2M device network scenario; solid line stands for a high-quality link and dashed line stands for a medium-quality link.

the links accordingly. Cooperation at the link layer is simpler to implement in duty-cycled systems compared to the more complex PHY cooperative schemes.

A cost metric similar to ours to characterise the link quality was previously investigated in [15]. However, in [15], indefinite packet retransmissions until success were assumed, which results in significant energy cost in outage conditions, thus diverging from the optimal solution. The cost metric in [15] was verified experimentally, while we take an analytical approach that is validated by comprehensive simulations. Authors in [16] study the problem of dynamic data forwarding depending on the link quality from the routing perspective. Based on the results, they conclude that the dynamic forwarding provides highly robust and reliable systems.

3. System Model

3.1. Scenario. An M2M device network is considered, consisting of N devices and a data collector denoted sink. A traffic pattern is convergecast towards the sink and the devices may opt for a direct or a multihop transmission to the sink if a direct link cannot be established. The number of hops on a multihop path is denoted as k. Any link that can improve progress to the sink by decreasing k is denoted as a direct link. To transmit a packet to the sink, various combinations of links between a pair of devices can be formed, as shown in Figure 1. The links represented with solid lines are short-range and on average more reliable than the medium-range links represented with dashed lines. Therefore, we classify a short-range link with small probability of error as a high-quality link, distinguished from a medium-range link with greater probability of error denoted as a medium-quality link. The unreliability that is bound to the medium-range links is the main reason why they are usually discarded, even though they offer better progress to the sink.

3.2. Medium Access Control Layer. We consider time synchronised channel hopping (TSCH) mode of the MAC protocol in IEEE 802.15.4e [3]. It defines a fixed time division multiple access (TDMA) frame structure that is centrally scheduled. Each *link* formed by a pair of neighbour devices is assigned to a unique time slot that repeats in a cyclical manner. The receiver wakes up only in the assigned slot and may enter a sleep state (i.e., switch off its radio transceiver) for the rest of time. After it has woken up, it either receives a packet if the transmitter has one to send or goes quickly back to sleep if a packet preamble is not detected within a short, predefined time interval, that is, a fraction of the slot duration. The slot without a packet transmission is denoted as the *idle listen* slot. We consider dedicated links that are scheduled for each transmitter-receiver pair to prevent packet collisions.

3.3. CDC-ARQ Overview. CDC-ARQ technique enables the efficient use of medium-range links. CDC-ARQ operates as follows: each new transmission is first attempted over the direct, medium-quality link, for example, over link 4-2 in Figure 1. If it fails, the packet is redirected to the multihop path that offers higher reliability (4-3-2). Time slots are assigned both for the medium-range and medium-quality links, as well as for the backup multihop path that consists of short-range equidistant links, as shown in Figure 2. If the initial attempt over medium-quality link succeeds, the receivers on backup links only perform idle listen in the fraction of their slots. With CDC-ARQ, two forwarding options cooperate to provide a better service for the current channel realisation. Short-range links are approximated to be equidistant and k times shorter than the direct links. The goal is to find a link distance coupled with the appropriate forwarding scheme that results in minimum energy consumption.

3.4. Modulation and Coding. To exemplify the analysis, we focus on the transceiver of the IEEE 802.15.4 radios, working in the 2.4 GHz band, whose bit error rate p_b is given by [17]

$$p_b(\gamma) = \frac{8}{15} \cdot \frac{1}{16} \cdot \sum_{i=2}^{16} -1^i \binom{16}{i} e^{(20\gamma(1/i-1))}, \quad (1)$$

where γ is the instantaneous signal-to-noise-ratio (SNR) at the receiver. The packet success probability p_s assuming independent bit errors is

$$p_s(\gamma) = (1 - p_b)^L, \quad (2)$$

where L is the packet size in bits. The packet error probability p_e is therefore $p_e = 1 - p_s$. Note that p_s depends on the instantaneous γ characterized by the channel dynamics.

3.5. Channel Model. The wireless signal strength decays exponentially with distance as described by the *large-scale* pathloss channel model. Superimposed on this deterministic value is the random *medium-scale* fading model (also known as *shadowing*) and *small-scale* multipath fading. Given the transmission power P_t, the mean power at the receiver P_r is therefore

$$P_r(\text{dBm}) = P_t(\text{dBm}) - \text{PL}(d_0) - 10\alpha \log_{10} d - X_\sigma, \quad (3)$$

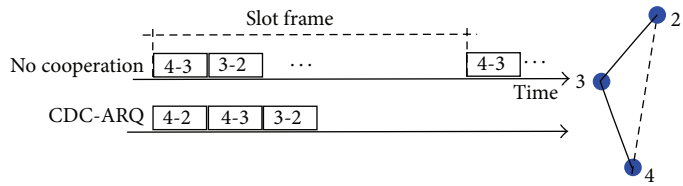


Figure 2: TDMA slot scheduling with and without cooperation.

where $\mathrm{PL}(d_0)$ is the pathloss at reference distance d_0, α is the pathloss exponent, d is the link distance from the transmitter, and $X_\sigma \sim N(0,\sigma)$ is a normal random variable (on dB scale). On a linear scale, normal distribution becomes lognormal such that X_σ models the random component of a wireless signal. Given the average receiver thermal noise P_n in dBm, the mean SNR at the receiver (when fading effects are averaged out) is $\mu(\mathrm{dB}) = P_r - P_n$. Recall that the instantaneous SNR at the receiver that varies in space and time is denoted as γ. For a given μ, the instantaneous γ expressed on a linear scale is lognormally distributed and conditioned on μ and with standard deviation σ, both expressed in dBs, the probability density function (pdf) of which is

$$p_\gamma(\gamma \mid \mu) = \frac{10}{\ln 10\sqrt{2\pi}\sigma\gamma} \exp\left[-\frac{(10\log_{10}\gamma - \mu)^2}{2\sigma^2}\right]. \quad (4)$$

We assume that channel hopping in the TSCH scheme alleviates small-scale multipath fading for packet retransmissions. The shadowing effect resulting from large moving obstacles remains. Similar channel model with lognormal shadowing was adopted in [18], but static environment was assumed where γ remains constant for long time intervals.

In dynamic environments, the assumption of γ being constant in time no longer holds. The empirical study of low-power wireless links in [5] showed that γ is correlated in time on scales larger than a packet duration, which results in effect denoted as link *burstiness*. In addition, medium success rate on a link is not the result of a corresponding constant packet success probability $p_s(\gamma)$, but rather the consequence of frequent oscillations between the nearly perfect and outage states that, when averaged, result in a medium-quality link. Based on these empirical observations, that are revisited and confirmed in [19], and essentially adopting a block-fading channel, we make the following assumptions:

(i) the realization of the channel fading, both comprising small-scale fading and shadowing, is drawn from distribution in (4);

(ii) the system is duty-cycled such that for every original packet a new realization of γ according to lognormal distribution is encountered;

(iii) in case of a transmission error, the packet is retransmitted up to $N_{\max}$ number of attempts (truncated-ARQ) where each retransmission encounters again the same value of γ on the same link.

The last assumption is based on the fact that packet retransmissions are sufficiently close in time to experience link burstiness described in [5]. New packets however are generated at a rate at which burstiness effect disappears. Therefore, independent channel realisation is assumed for new packets.

3.6. Energy Model. From the exemplary radio transceiver data sheets, for example, [20], we distinguish two basic radio power modes (voltage $V_{DD} = 3\,\mathrm{V}$), with the associated current consumption:

(1) *sleep*, $I_s \approx 0$,

(2) *awake* (also, *on*), that further exhibits two submodes:

(i) *active*, either transmitting or actively receiving $I_{tx} \approx I_{rx} = I_a = 20\,\mathrm{mA}$,

(ii) *idle* (*listen*), waiting for signal $I_{\mathrm{id}} = 2\,\mathrm{mA}$.

We assume that the output transmission power is set to $P_t = 0\,\mathrm{dBm}$. The energy is calculated as the product of the power consumption of a mode and the time spent in that power mode (power $P_x = V_{DD} \cdot I_x$, energy $E_x = P_x \cdot T$, where x can stand for sleep s, transmitting tx, actively receiving rx, or idle listen id). Power modes change within a slot according to TSCH algorithm. Before a transmission, clear channel assessment (CCA) is performed. There is no random backoff and no contention (recall that slots are dedicated). CCA is introduced for coexistence with other systems (e.g., IEEE 802.11 network) with whom the same physical space is shared. Therefore, a device performing CCA in TSCH mode listens to the wireless channel for a fixed time interval to determine whether it is free. If yes, a packet transmission follows; if not, a transmission is delayed until the next slot. An idle radio power mode is activated when there is no data transmission resulting in the idle listen slot of duration T_{id}, if a device is performing clear channel assessment for the time duration T_{cca} and between the end of data transmission and before the beginning of acknowledgement (ACK) for T_{ad} (ACK delay). Since the energy consumption in the sleep state can be neglected, we do not include it in our model. To calculate the time duration of active power modes, packet lengths (in bits) are divided with the data bit rate, $T_{\mathrm{data}} = L/R_b$ and $T_{\mathrm{ack}} = L_{\mathrm{ack}}/R_b$. Finally, we can define three basic energy consumption components associated with the MAC layer described in Section 3.2:

$$\begin{aligned} E_{\mathrm{data}} &= T_{\mathrm{cca}} \cdot P_{\mathrm{id}} + T_{\mathrm{data}} \cdot (P_{tx} + P_{rx}), \\ E_{\mathrm{ack}} &= T_{ad} \cdot P_{\mathrm{id}} + T_{\mathrm{ack}} \cdot (P_{tx} + P_{rx}), \\ E_{\mathrm{idle}} &= T_{\mathrm{id}} \cdot P_{\mathrm{id}}. \end{aligned} \quad (5)$$

Note that the energy components include the energy spent both at a transmitter and at a receiver.

4. Energy Analysis for Low-Power Links

4.1. Average Number of Transmissions per Packet. If the upper bound on the allowed number of transmission attempts per packet $N_{\max}$ was not set, huge amounts of energy would be wasted in periods of channel outage because the packet would be continuously retransmitted without success until the channel became available again. This is why the truncated-ARQ is applied to discard a packet if it was transmitted $N_{\max}$ times and still not successfully delivered. While truncated-ARQ can save significant amounts of energy, it produces a certain packet loss on a link that can be measured with the outage probability parameter.

Taking the channel model described in Section 3.5, the average number of transmissions for one realization of γ is then

$$\begin{aligned} N_{tx}(\gamma) = {} & 1 \cdot (1 - p_e(\gamma)) + 2 \cdot p_e(\gamma)(1 - p_e(\gamma)) \\ & + 3 \cdot p_e(\gamma)^2 (1 - p_e(\gamma)) + \cdots \\ & + N_{\max} p_e(\gamma)^{N_{\max}-1} (1 - p_e(\gamma)) \\ & + N_{\max} p_e(\gamma)^{N_{\max}}. \end{aligned} \tag{6}$$

Recall that γ remains constant for all retransmissions. The last member of the sum denotes the packets that were received erroneously $N_{\max}$ times and discarded. After some simple algebra we get

$$N_{tx}(\gamma) = \sum_{n=1}^{N_{\max}} p_e(\gamma)^{n-1} = \sum_{n=1}^{N_{\max}} (1 - p_s(\gamma))^{n-1}. \tag{7}$$

The average number of transmission attempts per packet, over all γ realization, is then

$$\overline{N}_{tx}(\mu) = \int_0^{\infty} N_{tx}(\gamma)\, p_\gamma(\gamma \mid \mu)\, d\gamma. \tag{8}$$

Note that because the integrand in (7) is applied instead of $p_s(\gamma)$ (the latter is usually studied), γ is kept constant for all the retransmissions to reflect the empirically observed link temporal correlation. No approximation for high SNR values can be applied since the lower limit of integration is zero. Therefore we solve the integral in (8) numerically.

4.2. Outage Probability. We define the outage probability p_{out} as the probability that a packet is discarded after $N_{\max}$ failed transmission attempts. To calculate p_{out}, we use the average number of errors associated to a realization of γ as follows:

$$\begin{aligned} N_{\text{err}}(\gamma) &= \sum_{n=0}^{\infty} n \cdot (p_e(\gamma))^n \cdot (1 - p_e(\gamma)) \\ &= \frac{p_e(\gamma)}{1 - p_e(\gamma)}. \end{aligned} \tag{9}$$

A link is in outage if the current value of γ is such that the associated average number of errors is $N_{\text{err}} \geq N_{\max}$. Taking the inverse function of (9), we can find the threshold γ_{th} such that $\gamma_{\text{th}} = N_{\text{err}}^{-1}(N_{\max})$. Therefore, if the current realization of γ satisfies $\gamma \leq \gamma_{\text{th}}$, the link is said to be in outage. If γ is lognormally distributed, there is a nonzero outage probability for every device involved in communication regardless of the value of μ. The outage probability can be computed as [4]

$$\begin{aligned} p_{\text{out}} &= \int_0^{\gamma_{\text{th}}} p_\gamma(\gamma \mid \mu)\, d\gamma \\ &= Q\left(\frac{\mu - 10 \log_{10}(\gamma_{\text{th}}(N_{\max}))}{\sigma} \right), \end{aligned} \tag{10}$$

where $Q(\cdot)$ is the tail integral of a unit Gaussian probability density function.

4.3. Energy Consumption Analysis. The mean energy spent both at the transmitter and at the receiver to exchange a packet over a link, that is, per time slot, is

$$\begin{aligned} E_{\text{link}} = {} & \overline{N}_{tx} \cdot E_{\text{data}} + (1 - p_{\text{out}}) \cdot E_{\text{ack}} \\ & + (N_{\max} - \overline{N}_{tx}) \cdot E_{\text{idle}}, \end{aligned} \tag{11}$$

where $\overline{N}_{tx}$ and p_{out} are defined in (8) and (10), respectively, and the energy components are defined in (5). In order to avoid buffering packets coming from the upper layer in the transmit queue, possible retransmissions need to be considered when the time frame is designed. Therefore some slots result in idle slots when there are no retransmissions and they are represented by the last component in (11). Both $\overline{N}_{tx}$ and p_{out} depend on the mean SNR μ whose relation to the link distance can be derived from (3).

For a fixed distance to the destination, a device may opt for a direct, multihop, or CDC-ARQ transmission. In case of a multihop transmission, we refer to the multihop path made of consecutive, equidistant links. A path begins with the source device, continues over $k - 1$ relays, and finally ends at the destination. The total mean energy spent to transmit a packet over a multihop path is

$$E_{\text{tot}} = \sum_{n=0}^{k-1} (1 - p_{\text{out}})^k \cdot E_{\text{link}}, \tag{12}$$

because only those packets that were delivered to a relay are forwarded over the next link and p_{out} is the outage probability of each individual link. For a direct transmission when $k = 1$, (12) gives $E_{\text{tot}} = E_{\text{link}}$. However, to single out the efficient part of the total consumed energy, we are interested in the energy spent per *delivered* packet, because if the packet does not reach the destination, the energy spent in a transmission attempt is wasted. The unit of effective energy consumption is thus *Joule per delivered packet*. The probability of discarding a packet on a multihop path is

$$p_{\text{out}}^t = 1 - (1 - p_{\text{out}})^k. \tag{13}$$

Therefore, we calculate the effective mean energy per delivered packet on a multihop path as follows:

$$E_{\text{eff}} = \frac{E_{\text{tot}}}{1 - p_{\text{out}}^{t}}. \tag{14}$$

In a direct transmission when $k = 1$, $E_{\text{eff}} = E_{\text{link}}/(1 - p_{\text{out}})$.

Finally, CDC-ARQ represents a combination of the two, as it alternates between the medium-range link and the multihop transmission depending on channel conditions. Recall that a transmission is first attempted over a direct link. If it results in error, the packet is immediately redirected because the probability of error for the subsequent attempts on the same link is high. If however the attempt over the direct link results in success, the time slots on a backup path remain idle. Therefore, the mean energy spent in a CDC-ARQ scheme for $k = 2$ backup multihop path is

$$E_{\text{carq}} = p_{\text{out}}^{c} \cdot \left(E_{\text{data}} + E_{\text{tot}}\left(k = 2\right)\right) + \left(1 - p_{\text{out}}^{c}\right) \cdot \left(E_{\text{data}} + E_{\text{ack}} + kN_{\max}E_{\text{idle}} + \left(N_{\max} - 1\right)E_{\text{idle}}\right), \tag{15}$$

where p_{out}^{c} is the probability of redirecting to a multihop path because of a direct link outage. The probability p_{out}^{c} corresponds to $N_{\text{err}} = 1$ (error on a direct link) and can be calculated from (10) for the given link distance by replacing $\gamma_{\text{th}} = N_{\text{err}}^{-1}(1)$. When the packet is redirected, we need to include the energy spent for the initial attempt that resulted in error. Otherwise, only the energy of the successful transmission with ACK is consumed plus the idle slots on a backup path. The effective mean energy per delivered packet is then

$$E_{\text{eff}}^{c}\left(k = 2\right) = \frac{E_{\text{carq}}}{1 - p_{\text{out}}^{c} \cdot p_{\text{out}}^{t}\left(k = 2\right)}, \tag{16}$$

where E_{eff}^{c} denotes the energy of the cooperative forwarding scheme as opposed to the energy of the fixed forwarding scheme E_{eff}.

Although the expressions for the network energy consumption in CDC-ARQ scheme for $k > 2$ are tractable, they can be quite complex because of multiple medium-range links that can be formed. That is why we chose to validate the model in the simulation with the simplest case of $k = 2$, which, in its turn, also validates the implementation in the simulator. Then we proceed with the evaluation in the simulator only for $k > 2$. This step is in line with the general tendency to validate the developed ns-3 code and thus support the trustworthiness of the simulation output. In addition, a quick access to complex scenarios justifies the use of a simulator.

For better readability, the notations used throughout the paper are summarized in notations section.

5. Performance Analysis

5.1. Implementation in ns-3 Simulator. ns-3 is an open-source, discrete-event network simulator written in C++. It consists of libraries for various technology models that implement the protocol interface and packet format (including the headers) by closely following the definitions of the corresponding standard. This approach enables a reliable simulation of the real system and facilitates the integration with a testbed. The final goal of a fully supported model is to enable each simulated device to run an entire protocol stack and generate the output trace that is (almost) indistinguishable from the output of a real device. Given that ns-3 is a network simulator, the unit of granularity is a packet, which contains both payload (in case of data packets) and a corresponding header. Each layer of a protocol stack executes its specified role; for example, the MAC layer at the transmitting side generates the MAC header, adds it to the payload received from the upper layer, and forwards a packet to the PHY layer where it is serialised and transmitted over the wireless channel as a set of bits.

The functionality required for this study has been implemented by modifying a model for the IEEE 802.15.4 standard, whose name is *lr-wpan* and whose source code can be downloaded from [21]. The MAC implementation available in the initial *lr-wpan* model supports the nonbeacon, mesh mode with all devices in the default idle listen state of the radio transceiver. This model has been significantly extended to include the duty-cycling operation of the devices, to support the energy consumption measurements, and to enable the path diversity necessary for the CDC-ARQ scheme. Firstly, duty-cycling has been implemented with TDMA as explained in Section 3.2. Contention during the CCA phase has been disabled and replaced with the simple CCA without random backoff as specified in [3]. Next, the existing log-distance path loss channel model has been extended with a realistic shadowing model described in Section 3.5 and defined with a standard deviation σ and a channel correlation in time. The CDC-ARQ functionality has been implemented by introducing dedicated tags in the packet header. Finally, the energy consumption module has been extended to subscribe to the changes in the state of a PHY radio transceiver in order to obtain the energy consumption directly from the device (i.e., from the simulated radio). The energy module is not aware of the wireless channel nor of the MAC layer scheme; therefore the two cannot interfere with the energy reading. Because of this design choice, an independent comparison with the theoretical model is provided.

5.2. Results

(1) Simulation Parameters. The default values of simulation parameters are given in Table 1. The maximum number of transmission attempts $N_{\max}$ is recommended in the standard, as well as the data bit rate that corresponds to the chosen frequency band; the modulation considered is offset quadrature phase-shift keying (O-QPSK) with additional direct sequence spread spectrum (DSSS), whose bit error rate is given in (1). The default header is generated in the simulation as specified in the standard [3].

(2) Link Metrics. The first step in the validation of the energy model is to show that the presented metrics, namely,

Table 1: Simulation parameters.

Parameter	Value
P_t	0 dBm
σ	4 dB
R_b	250 kbits/s
$N_{\max}$	4
L	27 bytes
freq. band	2.4 GHz

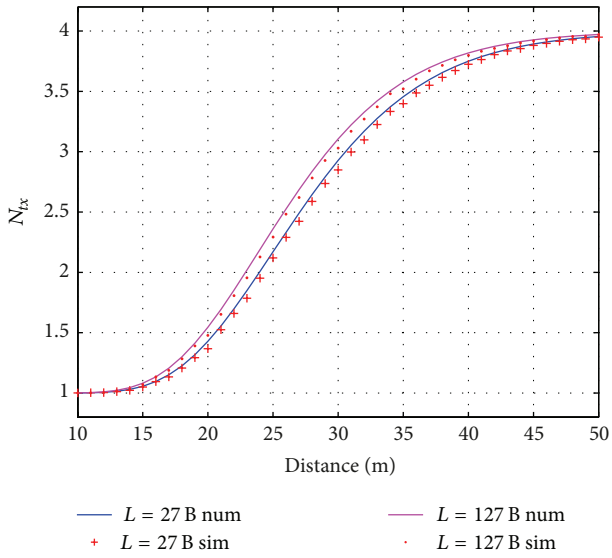


Figure 3: Average Number of Transmissions per Packet as a function of link distance; payload of minimum $L = 27$ bytes; and maximum length $L = 127$ bytes; truncated-ARQ, $N_{\max} = 4$.

$\overline{N}_{tx}$ paired with p_{out}, characterises a low-power wireless link reliably. To that aim, we observe a low-power link for various distances when the output transmission power P_t is fixed. Figure 3 shows the Average Number of Transmissions per Packet in a truncated-ARQ scheme over a low-power, shadowed wireless link. As the distance increases, more and more packets get discarded and $\overline{N}_{tx}$ approaches the maximum allowed number of attempts per packet. If the scheme was not truncated but indefinite retransmissions until successful delivery were permitted, in the considered shadowed environment the function $\overline{N}_{tx}$ would not be bounded. The unbounded function $\overline{N}_{tx}$ would tend to infinity which makes the analysis intractable. In general, system is designed such that outage states are brought to minimum because they destabilise the system. That is why the upper limit $N_{\max}$ had to be imposed on the number of transmission attempts. Not only does it limit the influence of outage states, but it also preserves energy and system resources, for the cost of some discarded packets. However, the probability of discarding a packet can be controlled either with careful system design or with CDC-ARQ. Figure 3 shows $\overline{N}_{tx}$ for the minimum and maximum packet lengths allowed by the standard (27 bytes $\leq L \leq$ 127 bytes). It can be seen that the packet size does not influence this metric significantly. The average number of

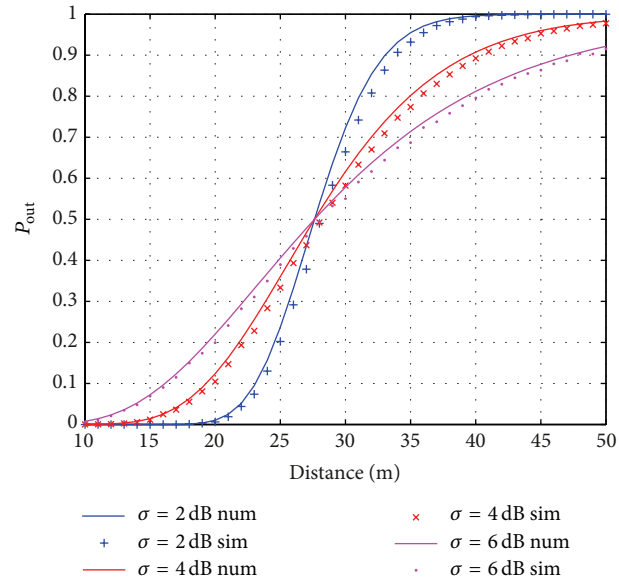


Figure 4: Outage probability as a function of link distance for different shadowing σ; output transmission power is set to 0 dBm.

transmissions is an important parameter in estimating the energy consumption since the energy is directly proportional to its value, as (11) shows. Note from Figure 3 that the results obtained in MATLAB for the numerical solution of (8) match well with the simulation results obtained in ns-3.

The (un)reliability of the scheme can be estimated with the number of packets that get discarded after $N_{\max}$ attempts. The outage probability p_{out} that complements the $\overline{N}_{tx}$ metric is shown in Figure 4 for several values of the standard deviation of shadowing σ. A value of the outage probability depicted in Figure 4 for a given link distance corresponds to a value of $\overline{N}_{tx}$ in Figure 3 for the same distance. For example, when $\overline{N}_{tx} = 2$ for $L = 27$ bytes, the $p_{\text{out}} = 0.3$ for $\sigma = 4$ dB; that is, 30% of packets are discarded on average over this link. With the increase in σ, the range of link distances with medium packet loss probabilities increases. In another words, without shadowing the graph would resemble the step function that produces the unrealistic circle coverage area with the perfect reliability within the circle and the absolute outage outside of it. The inclusion of shadowing effects makes every link unreliable, with the predictable (average) degree of unreliability p_{out}. The numerical estimation of p_{out} in (10) obtained in MATLAB agrees well with the number of discarded packets in ns-3 simulation. A slight disagreement is the consequence of the missed ACKs not considered in the theoretical model which cause retransmissions in the simulation. It has been determined by simulation that successfully delivered data packets whose ACK is not received do not exceed 5% of all the transmitted packets. The probability of unsuccessful ACK transmission approaches 5% for medium link outage probabilities, but it decreases to values below 2% for low or high outage probabilities. It can be seen in Figure 4 that the slight disagreement is highest precisely for the medium outage probabilities. In addition, observe that

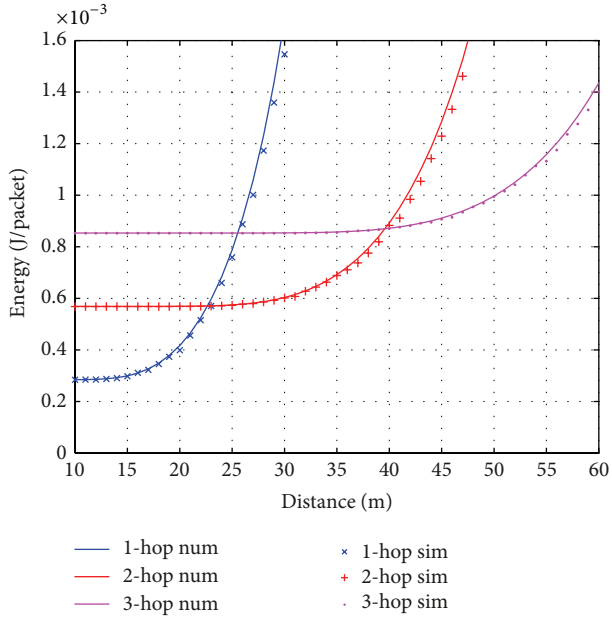


Figure 5: Total mean energy spent per delivered packet for direct, 2-hop, and 3-hop transmission over the total source-to-destination distance.

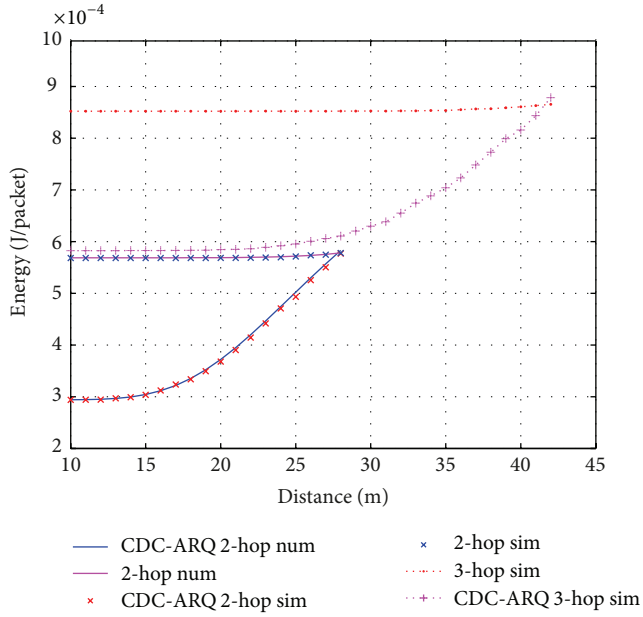


Figure 6: Total mean energy spent per delivered packet with the imposed outage constraint $p_{\text{out}} < 1\%$; multihop and CDC-ARQ forwarding compared.

a smaller σ yields smaller outage probability for $p_{\text{out}} < 0.5$ but provides worse delivery rates for $p_{\text{out}} > 0.5$. Nevertheless, in all studied cases, the direct transmission scheme is inefficient for larger distances and implies the use of either multihop transmission or a CDC-ARQ technique to avoid significant packet loss. We next determine for what distance range one of the three techniques is the most appropriate.

(3) Energy Consumption Optimisation. In the second step, energy consumption is measured in the simulation to verify the model derived in Section 4. While the first step focused on an isolated link, here we study various multinode scenarios. Results show how to optimise system's energy consumption in a multinode scenario. Given the reliability constraints, bounds for optimal link distances are shown to indicate the design choice between direct, multihop, or CDC-ARQ forwarding. For the strict reliability requirement, CDC-ARQ proves to be the most efficient forwarding strategy.

The mean consumed energy per delivered packet of the direct and multihop transmission is compared in Figure 5 for $k = 1$, $k = 2$, and $k = 3$ (the model for E_{eff} in (14) is valid for any k). For $k > 1$ the distance on the graph denotes the total distance traversed from source to destination via relays such that the individual link distance is k times smaller than the total distance. For distances $d < 23$ m there is no need for the multihop transmission because the direct link provides good service while consuming less energy. For $23\,\text{m} < d < 40\,\text{m}$ the optimal setting requires the transmission over two hops, that is, over two equidistant links of $d/2$ meters, past that distance over three hops, and so on. The threshold distance can be obtained numerically from (14) by substituting the corresponding k or from the simulation as Figure 5 shows. The energy consumption significantly increases as the outage probability for the corresponding link distance increases. Namely, packet loss presents a waste of energy resources and has considerable effect on the mean energy consumption. Note in Figure 4 that at distances close to the threshold link distances of $d = 23\,\text{m}$ ($k = 1$) and $d = 20\,\text{m}$ ($k = 2$), even 10–30% of packets can be lost. Therefore, the results presented in Figure 5 apply to systems with loosened reliability requirements but are not suitable for reliability-sensitive systems.

For the applications with strict reliability requirements, link distances must be decreased. In the design phase, maximum p_{out} is fixed in (10) to get the maximum link distance that meets the requirement. For example, when the link outage is set to $p_{\text{out}} < 1\%$, from (10) we get $d = 14\,\text{m}$ as the maximum link distance that still satisfies the outage constraint. Number of hops is next devised depending on the total distance between the source and the destination. In practice this is usually achieved with link estimators that select high-quality and therefore mostly short-range links. The alternative CDC-ARQ technique meets the same reliability requirement with lower energy consumption by including medium-quality links in the communication, besides backup high-quality links. To verify this hypothesis, we study the scenario shown in Figure 1. In this scenario, there is a two-hop path (4-3-2) that can alternate with a medium-quality link (4-2) and a three-hop path (7-6-5-1) that alternates with two medium-quality links (7-5 and 6-1).

Figure 6 shows how much energy can be saved by applying CDC-ARQ in comparison to using the fixed multihop path exclusively. For total distances $14\,\text{m} < d < 28\,\text{m}$ without CDC-ARQ on average 570 μJ of energy is spent per packet on a fixed two-hop path. CDC-ARQ that combines

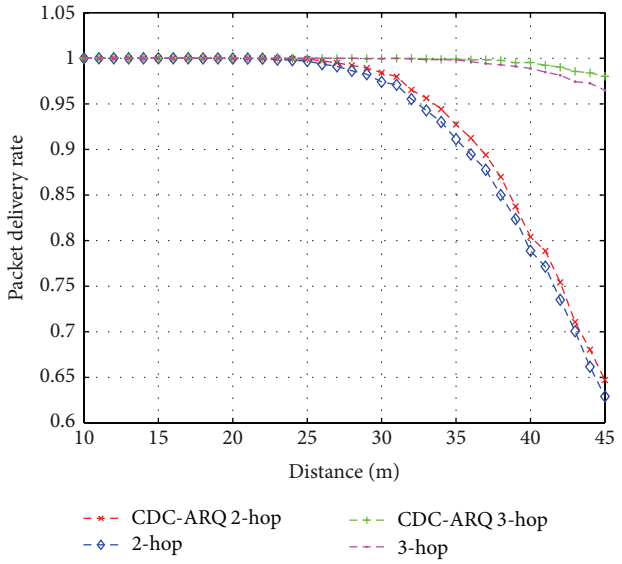


FIGURE 7: Total packet delivery rate for the given source-to-destination distance.

a medium-quality link up to 28 m in length with the two-hop path consumes less energy. The energy savings in CDC-ARQ are obtained when channel conditions allow successful transmissions over a medium-quality link and eventually take less hops to reach the destination. Failures on the medium-quality link represent the cost of the CDC-ARQ. As the link distance increases, there will be more failures and consequently more redirections to the two-hop path. It can be observed that the energy consumption increases accordingly until it becomes equal to the fixed multihop transmission. At the distance $d = 28$ m, about the half of packets are transmitted over a medium-quality link and the other half is redirected. This is a threshold point. Results obtained in simulation agree with the model in (14) and (16).

For source-to-destination distances $28\,\text{m} < d < 42\,\text{m}$, three short-range links are necessary to satisfy the outage constraint. Their mean energy consumption is 850 μJ per delivered packet. When medium-quality links participate in the communication, the energy consumption can again be significantly decreased as Figure 6 shows. Three-hop scenario was verified in simulation only given its analytical complexity. In conclusion, when the reliability demands require multihop transmission over short-range links, CDC-ARQ should be applied to decrease the mean energy consumption per delivered packet.

Although the data packet size L does not influence $\overline{N}_{tx}$ significantly, with the larger L, E_{data} in (5) increases. Consequently, the total energy spent to deliver a packet is larger; for example, $E_{\text{eff}} = 1.24$ mJ when $L = 127$ bytes, $k = 2$, and $p_{\text{out}} < 0.01$ compared to $E_{\text{eff}} = 570\,\mu$J when $L = 27$ bytes. However, apart from the absolute value, the characteristics of energy consumption behaviour for the permitted values of L ($27\,\text{bytes} < L < 127\,\text{bytes}$) closely resemble the ones already commented and shown in Figures 5 and 6.

(4) Packet Delivery Rate. Finally, in order to demonstrate that the presented results satisfy the outage constraint set in advance, the total packet delivery rate for the given source-to-destination distance is measured in the simulation and the results are shown in Figure 7. Just as the model predicts, when $k = 2$ and for the total distances up to $d = 28$ m, the number of discarded packets is limited to less than 1%. The same is confirmed for $k = 3$ and total distances up to $d = 42$ m. The delivery rate decreases for distances greater than the threshold distance obtained from the model such that the reliability requirement cannot be met any more. It can be observed in Figure 7 that CDC-ARQ performs slightly better than the fixed transmission over the multihop path.

6. Conclusion

In this paper, we introduced a metric adequate for low-power wireless links denoted as Average Number of Transmissions per Packet. The metric considers shadow fading and truncated-ARQ. Based on this metric, we were able to calculate the energy consumption of devices compatible with IEEE 802.15.4e, that is, suitable for duty-cycled, low-power, and energy-efficient M2M networks. The energy model presented in this work was validated in the ns-3 network simulator in the realistic simulation environment that mimics the functioning of an actual device. Based on the energy model, we determined the optimal operating regions of direct, multihop, and CDC-ARQ forwarding, as well as the implications of using each. All these results provide useful guidelines on the design of an energy-efficient network, for systems with either loose or strict reliability requirements.

The presented model and the simulation tools can be further refined to include channel hopping and the related channel model implications. Also, more realistic battery models could measure device lifetime. This will be investigated in our future work.

Notations

k:	Number of hops on a multihop path
d:	Link distance in meters
γ:	Instantaneous SNR
$p_s(\gamma)$:	Packet success probability
$p_e(\gamma)$:	Packet error probability
μ:	Mean SNR
σ:	Shadowing standard deviation
$N_{\max}$:	Maximum number of tx attempts per packet
$\overline{N}_{tx}(\mu)$:	Average number of tx per packet
$N_{\text{err}}(\gamma)$:	Number of tx errors
L:	Packet size
p_{out}:	Outage probability when $N_{\max} = 4$
p^c_{out}:	Outage probability when $N_{\text{err}} = 1$
E_{data}:	Energy consumed during data transmission
E_{eff}:	Effective mean energy per delivered packet.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been partially funded by the Spanish Government through the research project GEOCOM (TEC2011-27723-C02-01) and by the European Commission under the FP7 Program through the projects NEWCOM# (FP7-318306) and ADVANTAGE (MC-ITN-607774).

References

[1] M. R. Palattella, N. Accettura, X. Vilajosana et al., "Standardized protocol stack for the internet of (important) things," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1389–1406, 2013.

[2] L. Zhang, R. Ferrero, E. R. Sanchez, and M. Rebaudengo, "Performance analysis of reliable flooding in duty-cycle wireless sensor networks," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 2, pp. 183–198, 2014.

[3] IEEE Standard, "Low rate wireless personal area networks (LR-WPANs) amendment 1: MAC sublayer," IEEE Standard 802.15.4e, 2012.

[4] P. Mary, M. Dohler, J.-M. Gorce, G. Villemaud, and M. Arndt, "M-ary symbol error outage over Nakagami-m fading channels in shadowing environments," *IEEE Transactions on Communications*, vol. 57, no. 10, pp. 2876–2879, 2009.

[5] K. Srinivasan, P. Dutta, A. Tavakoli, and P. Levis, "An empirical study of low-power wireless," *ACM Transactions on Sensor Networks*, vol. 6, no. 2, article 16, pp. 1–49, 2010.

[6] M. Dianati, X. Ling, K. Naik, and X. Shen, "Performance analysis of the node cooperative ARQ scheme for wireless ad-hoc networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, pp. 3063–3067, St. Louis, Mo, USA, December 2005.

[7] T. Predojev, J. Alonso-Zarate, and M. Dohler, "Energy efficiency of cooperative ARQ strategies in low power networks," in *Proceedings of the 31st Annual IEEE Conference on Computer Communications Workshops (IEEE INFOCOM '12)*, pp. 139–144, Orlando, Fla, USA, March 2012.

[8] T. Predojev, J. Alonso-Zarate, and M. Dohler, "Energy evaluation of a cooperative and duty-cycled ARQ scheme for Machine-to-Machine communications with shadowed links," in *Proceedings of the 24th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC '13)*, pp. 1640–1644, London, UK, 2013.

[9] Network simulator 3, http://www.nsnam.org/.

[10] J. Deng, Y. S. Han, P.-N. Chen, and P. K. Varshney, "Optimal transmission range for wireless ad hoc networks based on energy efficiency," *IEEE Transactions on Communications*, vol. 55, no. 9, pp. 1772–1782, 2007.

[11] K. Vardhe, C. Zhou, and D. Reynolds, "Energy efficiency analysis of multistage cooperation in sensor networks," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, Miami, Fla, USA, December 2010.

[12] G. de Oliveira Brante, M. Kakitani, and R. Souza, "Energy efficiency analysis of some cooperative and non-cooperative transmission schemes in wireless sensor networks," *IEEE Transactions on Communications*, vol. 59, no. 10, pp. 2671–2677, 2011.

[13] J. Alonso-Zarate, L. Alonso, and C. Verikoukis, "Performance analysis of a persistent relay carrier sensing multiple access protocol," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5827–5831, 2009.

[14] M. H. Alizai, O. Landsiedel, J. A. B. Link, S. Gotz, and K. Wehrle, "Bursty traffic over bursty links," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09)*, pp. 71–84, Berkeley, Calif, USA, November 2009.

[15] D. Lal, A. Manjeshwar, F. Herrmann, E. Uysal-Biyikoglu, and A. Keshavarzian, "Measurement and characterization of link quality metrics in energy constrained wireless sensor networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 1, pp. 446–452, San Francisco, Calif, USA, December 2003.

[16] T. Iwao, K. Yamada, M. Yura et al., "Dynamic data forwarding in wireless mesh networks," in *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm '10)*, pp. 385–390, Gaithersburg, Md, USA, 2010.

[17] IEEE Standard, "Low rate wireless personal area networks (LR-WPANs)," IEEE Standard 802.15.4, 2011.

[18] M. Z. Zamalloa and B. Krishnamachari, "An analysis of unreliability and asymmetry in low-power wireless links," *ACM Transactions on Sensor Networks*, vol. 3, no. 2, Article ID 1240227, 2007.

[19] N. Baccour, A. Koubaa, L. Mottola et al., "Radio link quality estimation in wireless sensor networks: a survey," *ACM Transactions on Sensor Networks*, vol. 8, no. 4, article 34, 2012.

[20] CC2430 datasheet, 2010, http://www.ti.com/product/CC2430.

[21] Low-rate, wireless personal area network (LR-WPAN) ns-3 model, http://code.nsnam.org/tomh/ns-3-lr-wpan/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 341309, 12 pages
http://dx.doi.org/10.1155/2014/341309

Research Article

Game-Theoretic Based Distributed Scheduling Algorithms for Minimum Coverage Breach in Directional Sensor Networks

Jin Li,[1,2] Kun Yue,[2,3] Weiyi Liu,[3] and Qing Liu[1,2]

[1] *School of Software, Yunnan University, Kunming 650091, China*
[2] *Key Laboratory of Software Engineering of Yunnan Province, Kunming 650091, China*
[3] *School of Information Science and Engineering, Yunnan University, Kunming 650091, China*

Correspondence should be addressed to Jin Li; ljatynu@gmail.com

Received 27 September 2013; Revised 23 January 2014; Accepted 17 April 2014; Published 25 May 2014

Academic Editor: Christos Verikoukis

A directional sensor network, where a lot of sensors are intensively and randomly deployed, is able to enhance coverage performances, since working directions can be partitioned into different K covers which are activated in a round-robin fashion. In this paper, we consider the problem of direction set K-Cover for minimum coverage breach in directional sensor networks. First, we formulate the problem as a game called direction scheduling game (DSG), which we prove as a potential game. Thus, the existence of pure Nash equilibria can be guaranteed, and the optimal coverage is a pure Nash equilibrium, since the potential function of DSGs is consistent with the coverage objective function of the underlying network. Second, we propose the synchronous and asynchronous game-theoretic based distributed scheduling algorithms, which we prove to converge to pure Nash equilibria. Third, we present the explicit bounds on the coverage performance of the proposed algorithms by theoretical analysis of the algorithms' coverage performance. Finally, we show experimental results and conclude that the Nash equilibria can provide a near-optimal and well-balanced solution.

1. Introduction

In recent years, wireless sensor networks (WSNs) have attracted much attention as the promising platform for many applications, such as environmental monitoring and battlefield surveillance [1]. As a fundamental problem for WSNs, coverage optimization has been explored thoroughly in networks based on an omnidirectional sensing model [2]. Recently, with the advent and introduction of video sensors, ultrasonic sensors, and infrared sensors, coverage control algorithms for directional wireless sensors networks (dWSNs) have become an active subject and the state of the art is well surveyed in [3].

Power conversation is still a critical issue in dWSNs since the directional sensors in the network are usually battery-powered and nonrechargeable devices. Therefore, designing coverage optimization algorithms with energy efficiency is quite challenging for successful applications of dWSNs. One approach to meeting these challenges is to partition the working directions of sensors into K covers. By activating a different cover at each time slot and cyclically shifting through these covers, thereby, the network's lifetime can be extended effectively by a factor of K [4].

Some efforts have recently been devoted to the research of coverage optimization with energy efficiency for directional sensor networks. For example, Ai and Abouzeid [5] proposed a directional sensing model, where a sensor is allowed to work in several directions, and the objective is to find a minimal set of directions that can cover all targets. Cai et al. [6] defined the multiple directional cover set problem of organizing the directions of sensors into a group of nondisjoint cover sets to extend the network lifetime, in order to maximize the network lifetime of a directional sensor network. The network lifetime is defined as the time duration when each target is covered by the working direction of at least one active sensor. Following the work in [6], Wen et al. [7] gave the method for prolonging the lifetime of networks based on the combination of equitable direction optimization algorithm and the neighbors sensing scheduling protocol.

Generally, the aforementioned work is mainly to maximize the network lifetime of directional sensor networks while covering all targets, which is quite strict for general coverage problems. Sometimes, due to energy constraint, coverage breach [8] (i.e., targets are not covered) may occur if available working directions in a cover are not enough to cover all the targets. Instead of finding the maximum number of directional covers for complete target coverage, the problem of direction set K-Cover for minimum coverage breach (dKC-MCB) is reduced to schedule the working directions into K covers (K is predefined) to minimize the coverage breach. Although considerable research work has been devoted to the problem of set K-Cover for minimum coverage breach in omnidirectional sensing model [4, 9, 10], to our knowledge, few research efforts have been devoted to the problem of dKC-MCB. Even Yang et al. [11] dealt with the problem of minimum coverage breach under lifetime constraints in directional sensor network by formulating the problem as an integer programming and solving the problem by centralized greedy algorithms.

Although the existing research works have achieved some success on coverage optimization with energy efficiency in directional sensor network, some challenges still remain unanswered, especially for the problem of dKC-MCB. As mentioned in the paper [11], since the directional sensors are energy constrained, distributed coverage optimization algorithms need to be exploited where a sensor takes coverage optimization decisions independently, based purely on communications with its neighbors.

Game theory [12] is a mathematical theory about modeling and analyzing the strategic interactions among intelligent, rational decision makers. Recently, game theory is beginning to emerge as a powerful tool for the design optimization algorithms that can be distributed across many decision makers [13]. The core advantages of game theory for distributed optimization lie in that it provides a hierarchical decomposition between the distribution of the optimization problem (game design) and the specific local decision rules (distributed algorithms). Particularly, if the game is designed as a potential game [14], then there is a possibility that local decision dynamics can achieve convergence to pure Nash equilibrium which coincides with a desirable outcome of the original optimization problem.

Inspired by the previous discussion, in this paper, the problem of dKC-MCB is formulated as a game. Two game-theoretic based distributed algorithms are proposed to solve the problem. Specifically, the principal contributions of this paper are as follows.

(1) We first formulate dKC-MCB as a game: directions scheduling game (DSG). Sensors, as players of the game, interact with each other. Each sensor makes decisions independently to maximize its individual coverage utility. The utility for a sensor is defined as the sum of the marginal coverage contribution of working directions to networks coverage.

(2) We then prove that DSG is a potential game, whose potential function is the same as the optimization objective function of dKC-MCB. This correspondingly enables the design of a coverage optimization scheme that induces the equilibrium of a DSG consistence with the optimal coverage of the underlying network objective. Moreover, this consistency allows us to establish near-optimal performance for sensor dynamics applied to the original network coverage optimization, since the natural sensors' utility-update dynamics converge to a Nash equilibrium.

(3) We propose the synchronous and asynchronous distributed scheduling algorithms, which are proved to converge to pure Nash equilibria. Further, we analyze the coverage performance of the distributed algorithms from the theoretic perspective. We then present the explicit bounds on the coverage performance of both synchronous and asynchronous distributed algorithms.

2. Preliminaries

Game theory [12] is a mathematical tool that analyzes the strategic interactions among rational decision makers. Three major components in a strategic-form game model $G = \langle N, (A_i)_{i\in N}, (u_i)_{i\in N}\rangle$ are as follows.

(1) N is a finite set of n players.

(2) $\mathbf{A} = A_1 \times A_2 \times \cdots \times A_n$, where A_i is a finite set of actions (or *pure strategies*) available to player i. For any set X, let $\Pi(X)$ be the set of all probability distributions over X. Then the set of *mixed strategies* for player i is $s_i \in S_i = \Pi(A_i)$. A mixed-strategy profile $\mathbf{s} \in S_1 \times \cdots \times S_n$ is Cartesian product of the individual mixed strategy. A vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbf{A}$ is called a pure-strategy profile, often denoted by $\mathbf{a} = (a_i, \mathbf{a}_{-i})$, where a_i is a strategy of player i and $\mathbf{a}_{-i}$ is the strategy vector of other n-1 players.

(3) $u_i(\mathbf{a}) : \mathbf{a} \mapsto \Re$ is a real value utility function of player i. The utility function $u_i(\mathbf{a})$ measures the outcome of a player at a profile $\mathbf{a}$. In a game, a player chooses the proper actions against other players to maximize its individual utility.

A Nash equilibrium (NE) is a stable strategic profile where no player gets any incentive to unilaterally deviate from this profile. Thus, a Nash equilibrium in some sense is a reasonable outcome of a game. Following, we present the definition of a Nash equilibrium based on the definitions of the players' best response.

Definition 1 (see [12]). A best response for a player i to a mixed-strategy profile $\mathbf{s}_{-i}$ is a mixed strategy $s^* \in S_i$ such that $u_i(s_i^*, \mathbf{s}_{-i}) \geq u_i(s_i, \mathbf{s}_{-i})$ for all strategies $s_i \in S_i$.

Definition 2 (see [12]). A best response for a player i to a pure-strategy profile $\mathbf{a}_{-i}$ is a pure strategy $a^* \in A_i$ such that $u_i(a_i^*, \mathbf{a}_{-i}) \geq u_i(a_i, \mathbf{a}_{-i})$ for all strategies $a_i \in A_i$.

Definition 3 (see [12]). A mixed-strategy profile $\mathbf{s} = (s_1, \ldots, s_i, \ldots, s_n)$ is a Nash equilibrium, if, for all $i \in N$, s_i is a best response to the profile $\mathbf{s}$.

Definition 4 (see [12]). A pure-strategy profile $\mathbf{a} = (a_1, \ldots, a_i, \ldots, a_n)$ is a Nash Equilibrium, if, for all $i \in N$, a_i is a best response to the profile $\mathbf{a}$.

With respect to the existence of Nash equilibria of a game, Nash [12] proved that every game with a countable number of players and strategy set at least has a mixed strategy Nash equilibrium. However, in general, mixed Nash equilibrium only implies stable probability distributions over profiles, not the fixed play of a particular joint action profile. This type of uncertainty is unacceptable in many applications, such as our direction scheduling scenario. Instead, in this paper, we focused on the game with pure Nash equilibria. However, pure Nash equilibria are not available in every game. Recently, potential games, which were introduced by Monderer and Shapley [14], received increasing attention since they possess some desirable properties for engineering application scenario [15, 16], such as admitting a pure-strategy NE and best response dynamics can achieve the convergence to a pure-strategy NE.

Definition 5 (see [14]). A game $G = \langle N, (A_i)_{i\in N}, (u_i)_{i\in N}\rangle$ is a potential game if there exists a potential function $\Phi : \mathbf{A} \mapsto \Re$ such that, for for all $i \in N$, all $\mathbf{a}_{-i} \in \mathbf{A}_{-i}$ and $a_i, a_i' \in A_i$:

$$u_i\left(a_i', \mathbf{a}_{-i}\right) - u_i\left(a_i, \mathbf{a}_{-i}\right) = \Phi\left(a_i', \mathbf{a}_{-i}\right) - \Phi\left(a_i, \mathbf{a}_{-i}\right). \quad (1)$$

In a potential game, the change in a player's payoff that results from a unilateral change in strategy equals the change in the potential function. When formula (1) is satisfied, the game is called a potential game with the potential function $\Phi(\cdot)$. In a potential game, the set of pure Nash equilibria can be found by locating the local optima of the potential function, since the incentives of all players are mapped into one function. Thus, we have the following theorem.

Theorem 6 (see [12]). *Every potential game with finite strategies space at least has a pure-strategy Nash equilibrium [14].*

In addition to the existence of Nash equilibria, the quality of NE also needs to be considered. The concept of price of anarchy (PoA) [17] was introduced to measure the quality of Nash equilibria. Formally, let $\gamma(\mathbf{a}) : \mathbf{a} \mapsto \Re$ be the social objective function and let $\mathbf{a}^{OPT}$ be a social optimum if and only if $\mathbf{a}^{OPT} = \arg\max_{\mathbf{a}\in\mathbf{A}}\gamma(\mathbf{a})$.

Definition 7 (see [17]). The price of anarchy of a game G is defined as $\mathrm{PoA}(G) = \max_{\mathbf{a}\in\mathbf{A}}(\gamma(\mathbf{a}^{\mathrm{OPT}})/\gamma(\mathbf{a}))$.

Intuitively, the PoA of a game is the ratio of the objective social value of the worst possible Nash equilibrium to the value of the social optimum.

3. Problem Statement

In this section, we formally give the definitions of the problem of direction set K-Cover for minimum coverage breach in DSNs. We consider a directional sensor network of n directional sensors and m targets. Let $S = \{s_1, \ldots, s_n\}$ and $O = \{o_1, \ldots, o_m\}$ denote the directional sensors and the target set, respectively. Let $D = \{D_1, \ldots, D_n\}$ be the set of directions of all sensors. Each sensor s_i has a set of directions $D_i = \{d_{i,j} \mid j = 1, 2, \ldots, w\}$. Without loss of generality, l_i is the initial lifetime for each sensor s_i, which is the time duration when the sensor is in the active state all the time. t_k is the kth time slot of sensor networks and $\mathrm{TL} = \sum_{k=1}^{K} t_k$ is the total lifetime of sensor networks. Being similar to the problem of set K-Cover for minimum coverage breach defined in [2], the problem of dKC-MCB is formally identified as follows.

Definition 8. A direction schedule of directional sensor networks is a set of ordered pairs, (A_k, t_k), $k = 1, 2, \ldots, K$, in which $A_k \subset D$ is the set of working directions in time slot t_k. At a time slot t_k, any sensor s_i has at most one working direction; that is, for all $i, k, |D_i \cap A_k| \leq 1$. Since the lifetime for each sensor is limited, for any sensor s_i, one always has

$$\sum_{k=1}^{K} z_{i,k} \cdot t_k \leq l_i, \quad \begin{cases} z_{i,k} = 1, & \text{if } \left|D_i \cap A_k\right| = 1 \\ z_{i,k} = 0, & \text{if } \left|D_i \cap A_k\right| = 0. \end{cases} \quad (2)$$

Definition 9. Assume that (A_k, t_k), $k = 1, 2, \ldots, K$ is a direction schedule. The total network lifetime is given by $\mathrm{TL} = \sum_{k=1}^{K} t_k$. Let $F(A_k)$ be the set of coverage targets covered by the set of active directions A_k in time slot t_k. O is a target set. Total coverage breach is defined as $\sum_{k=1}^{K} t_k(|O| - |F(A_k)|)$. An average coverage breach rate is defined as $\sum_{k=1}^{K} t_k(|O| - |F(A_k)|)/\sum_{k=1}^{K} t_k$.

Based on the definitions of direction schedule and coverage breach, we give the definition of dKC-MCB.

Definition 10. Give a positive integer K and network lifetime threshold $T_{\mathrm{th}} > 0$, and construct a direction schedule (A_k, t_k), $k = 1, 2, \ldots, K$, such that the network lifetime is not less than T_{th}; that is $\sum_{k=1}^{K} t_k \geq T_{\mathrm{th}}$ and total coverage breach $\sum_{k=1}^{K} t_k(|O| - |F(A_k)|)$ is minimized. Equivalently, the problem of direction set K-Cover for minimum coverage breach is also defined to maximize $\sum_{k=1}^{K} t_k|F(A_k)|$.

Based on the results in [11], we know that the problem of dKC-MCB is NP-complete.

4. Direction Scheduling Game

A direction scheduling game is a triple $\mathrm{DSG} = \langle S, (A_i)_{s_i\in S}, (u_i)_{s_i\in S}\rangle$. S is a set of players, in which a player $s_i \in S$ is a directional sensor. A_i is the set of direction scheduling strategies of s_i. A strategy $a_i \in A_i$ is an allocation of working directions of $s_i \in S$ among the time slots and can be described as a set of ordered pairs $a_i = \{\langle d_{i,j}, t_k\rangle \mid d_{i,j} \in D_i, k = 1, 2, \ldots, K\}$. Due to the limitation of energy, a feasible strategy a_i should satisfy the next properties. (1) For all $\langle d_{i,j} \neq \emptyset, t_k\rangle \in a_i$, $\sum_{k=1}^{K} t_k \leq l_i$; that is, whole working lifetime of the sensor s_i is less than l_i. (2) At a time slot t_k, there is at most a working direction of s_i. At a time, direction scheduling strategies of all sensors are composed of a direction scheduling profile, which is denoted

by $\mathbf{a} = (a_1, \ldots, a_i, \ldots, a_n)$. At the profile $\mathbf{a}$, the coverage utility of a sensor s_i is defined as follows.

Definition 11. At a profile $\mathbf{a}$ and a time slot t_k, the working direction of sensor s_i is denoted by $d_{i,k}^{\mathbf{a}} \in D_i$. If there is no working direction for s_i at time slot t_k, $d_{i,k}^{\mathbf{a}} = \emptyset$. Given a profile $\mathbf{a}$, the coverage utility of a sensor s_i is denoted by $u_i(\mathbf{a})$ and then defined as follows:

$$u_i(\mathbf{a}) = \sum_{k=1}^{K} \left(F\left(A_k^{\mathbf{a}}\right) - F\left(A_k^{\mathbf{a}} \setminus \left\{d_{i,k}^{\mathbf{a}}\right\}\right)\right). \tag{3}$$

Intuitively, the utility function is defined as the sum of marginal coverage contribution of the sensor's working directions to networks coverage. In a direction scheduling game, a sensor tries to activate its directions in time slots where the sensor obtains most marginal coverage contribution. Obviously, sensors interacted with each other by maximizing each individual utility. Thus, the game is actually a dynamical interaction process. Would the interaction dynamics finally terminate and converge to a pure Nash equilibrium? In order to answer this question, we discuss the mathematical properties of direction scheduling games.

In a direction scheduling game, the set of working directions A_k in a time slot t_k is determined by the profile $\mathbf{a}$. Thus, at a profile $\mathbf{a}$, the set of working directions in a time slot t_k is denoted by $A_k^{\mathbf{a}}$ and the objective function of the problem of dKC-MCB is denoted by $Z(\mathbf{a}) = \sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|$. In what follows, we prove that direction scheduling game is a class of potential games, the potential function of which is $Z(\mathbf{a})$, that is, the objective functions of dKC-MCB.

Theorem 12. *A direction scheduling game is a potential game with the potential function* $\Phi(\mathbf{a}) = Z(\mathbf{a})$.

Theorem 13. *A pure Nash equilibrium of direction scheduling game is a local optimal solution of the objective function* $Z(\mathbf{a}) = \sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|$.

Theorem 14. *An optimal solution of the problem of dKC-MCB is a pure Nash equilibrium of a direction scheduling game.*

We give the proofs of Theorems 12, 13, and 14 in Appendices A.1, A.2, and A.3, respectively.

Some interesting mathematical properties for DSGs to solve the problem of dKC-MCB are well established by the results of Theorems 12, 13, and 14. Specifically, a DSG is proved to be a class of potential games by Theorem 12. From the previous result of Theorem 6, a DSG at least admits a pure Nash equilibrium. The connection between the solutions of dKC-MCB and pure equilibria of DSGs are described by Theorems 13 and 14. In particular, the consistence between the potential function and the objective function of dKC-MCB allows us to establish a near-optimal performance for local decision dynamics applied to the original network coverage optimization.

5. Distributed Scheduling Algorithms for Minimum Coverage Breach

In this section, based on the aforementioned properties of DSGs, we propose both synchronous and asynchronous distributed scheduling algorithms for the problem of dKC-MCB. From the perspective of game theory, both synchronous and asynchronous algorithms are actually some kind of best response dynamics of DSGs.

Specifically, n sensors are assumed to be randomly deployed in a target area. A sensor is supposed to know its location and be aware of the location of its neighbors through local communications. A sensor has a sensing range R_s and a communication range R_c. In this paper, we assume the communication range of a sensor node is at least as twice as the sensing range; that is, $R_c \geq 2R_s$. Denote S_i^u as the neighbors of a sensor s_i. In particular, for a sensor $s_j \in S_i^u$, the distance between s_j and s_i is less than $2R_s$. Actually, the utility of a sensor s_i depends on the strategies of the sensors within S_i^u. In other words, s_i obtains the sum of marginal coverage contribution only through local communications with the sensors within S_i^u. Thus, both synchronous and asynchronous algorithms are based on local information.

In a synchronous distributed algorithm, at each time step, sensors are considered to be able to synchronize their actions with one another according to a system clock. We also propose an asynchronous distributed algorithm for the case that maintaining tight clock synchronization is sometimes difficult. In asynchronous distributed algorithm, each sensor maintains its individual time clock. At a time step, only one sensor has an opportunity to update its direction scheduling strategy.

5.1. Synchronous Distributed Scheduling Algorithm. At each time step of the synchronous distributed scheduling algorithm (SDA), all the sensors are assumed to be able to synchronize their actions with one another according to a system clock. The algorithm will terminate based on a mark $END = \neg(Update_1 \vee \cdots \vee Update_n)$.

Definition 15. Let $Update_i$ be the mark of strategy update for a sensor s_i. If a sensor s_i is able to increase utility by updating its strategy, $Update_i$ is set to be true. Otherwise, $Update_i$ is set to be false. Thus, the termination mark of the algorithm is defined as the following Boolean symbol $END = \neg(Update_1 \vee \cdots \vee Update_n)$.

Specifically, if all the sensors send a message of update false to the system, the algorithm terminates. The synchronous distributed algorithm is shown in Algorithm 1.

Theorem 16. *A synchronous distributed scheduling algorithm converges to a pure Nash equilibrium.*

Proof. First, we prove that, at a time step of Algorithm 1, if more than one sensor is able to increase utility by updating their strategies, these sensors should be independent of one another on utility.

Input: An initial strategy of s_i; A system time clock $t = 0$; The mark of strategy update $Update_i \leftarrow true$.
Output: A Nash equilibrium strategy of $s_i : a_i^*$.
(1) **WHILE** $END = false$ **DO**
(2) Communicate with $\forall s_j \in S_i^u$ to obtain the strategy of s_j;
(3) Based on the definition of utility function, compute:
(4) $\Delta_i(\mathbf{a}(t)) = \max_{a_i'} \left(u_i(a_i', \mathbf{a}(t)_{-i}) - u_i(a_i, \mathbf{a}(t)_{-i})\right)$
(5) and $br_i(\mathbf{a}(t)) = \arg\max_{a_i'} \left(u_i(a_i', \mathbf{a}(t)_{-i}) - u_i(a_i, \mathbf{a}(t)_{-i})\right)$;
(6) **IF** $\Delta_i(\mathbf{a}(t)) > 0$ **THEN**
(7) Broadcasting $\Delta_i(\mathbf{a}(t))$ to $\forall s_j \in S_i^u$;
(8) **IF** $\Delta_i(\mathbf{a}(t)) > \max\left\{\Delta_j(\mathbf{a}(t)) \mid \forall s_j \in S_i^u\right\}$ **THEN**
(9) $a_i(t) \leftarrow br_i(\mathbf{a}(t))$; Sending $Update_i \leftarrow true$ to the system;
(10) **END IF**
(11) **ELSE**
(12) Broadcasting $\Delta_i(\mathbf{a}(t)) = 0$ to $\forall s_j \in S_i^u$; Sending $Update_i \leftarrow false$ to the system;
(13) **END IF**
(14) **END WHILE**

Algorithm 1: Synchronous distributed algorithm (SDA).

Let $\mathbf{a}(t)$ be the profile in a time step t. Denote the maximal increment of utility of sensor s_i at $\mathbf{a}(t)$ by $\Delta_i(\mathbf{a}(t))$; $\Delta_i(\mathbf{a}(t)) = \max_{a_i'}(u_i(a_i', \mathbf{a}(t)_{-i}) - u_i(a_i, \mathbf{a}(t)_{-i})) \geq 0$. When Algorithm 1 is not terminated, *END* is false. From the definition of *END*, there is at least one sensor updating its strategy. Assume that more than two sensors update strategies at the time step. From line 6 to 10 of Algorithm 1, two utility-dependent sensors are not able to update strategies at the same time step. Actually only the sensor of the highest increments on utility has an opportunity to update a strategy. In other words, when SDA algorithm proceeds, if more than one sensor updates strategies simultaneously, then these sensors should be independent of one another on coverage utility.

Second, we prove that when Algorithm 1 proceeds, the network coverage monotonically increases.

Since direction scheduling game is a potential game, as described in (1), the change in a sensor's utility that results from a unilateral change at a profile equals the change in a global potential function. Moreover, based on Theorem 12, the potential function of direction scheduling game is the same as the optimization objective function; that is, $\Phi(\mathbf{a}) = Z(\mathbf{a})$. Therefore, the increment in a sensor's utility that results from a unilateral change at a profile equals the increment in a network coverage.

If Algorithm 1 is not terminated, $\exists s_i \in S$, $\Delta_i(\mathbf{a}(t)) > 0$. Based on (1) and $\Phi(\mathbf{a}) = Z(\mathbf{a})$, $\Delta_i(\mathbf{a}(t)) > 0$ results in the increase of $\Phi(\mathbf{a}) = Z(\mathbf{a}) = \sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|$. Thus, network coverage *monotonically increases*. When there is more than one sensor that has opportunity to increase utility by updating strategies, from results of the first part of this proof, these sensors should be independent of one another on coverage utility. Thus, the increment of $Z(\mathbf{a})$ equals the sum of increment on utility resulting from the sensors updating strategies. Also, in this case, network coverage *monotonically increases*.

Since $Z(\mathbf{a})$ is *finite*, Algorithm 1 finally converges. At the time, no sensors are able to update strategies to increase utility; a pure Nash equilibrium is achieved. □

5.2. Asynchronous Distributed Scheduling Algorithm. In the asynchronous distributed scheduling algorithm (ADA), we use the asynchronous time model [18], which is well matched to the distributed nature of sensor networks. In particular, each sensor s_i has an independent clock ck_i whose "ticks" are distributed as a rate 1 Poisson process. A mark of updating strategy UT_i is set for each sensor s_i to permit updating its strategy. UT_i is set to be true when ck_i ticks and to be false at other time. The asynchronous distributed algorithm is shown in Algorithm 2.

Theorem 17. *An asynchronous distributed scheduling algorithm converges to a pure Nash equilibrium.*

Proof. Assume that n sensors are deployed in a target area. Each sensor s_i has an independent clock ck_i whose "ticks" are distributed as a rate 1 Poisson process. s_i has an opportunity to update its strategy when the mark UT_i is true. Since tick times are exponentially distributed, independent among sensors and independent across time, the tick time model can be equivalently formulated in terms of a single global clock ticking according to a rate n Poisson process. By letting $\{T_k\}_{k\geq 0}$ denote the arrival times for this global clock, then the individual clocks can be generated from the global clock by randomly assigning each to the sensors according to a uniform distribution. Based on properties of the Poisson process, at each arrival time of $\{T_k\}_{k\geq 0}$, there is only a sensor that has an opportunity to update its strategy.

Let $\mathbf{a}(t)$ be the profile in a time step t. Denote $\Delta_i(\mathbf{a}(t))$ as the maximal increment of utility of sensor s_i at $\mathbf{a}(t)$; that is, $\Delta_i(\mathbf{a}(t)) = \max_{a_i'}(u_i(a_i', \mathbf{a}(t)_{-i}) - u_i(a_i, \mathbf{a}(t)_{-i})) \geq 0$.

Input: An initial strategy of s_i; $UT_i \leftarrow false$; The mark of strategy update $Update_i \leftarrow true$.
Output: A Nash equilibrium strategy of $s_i : a_i^*$.
(1) **WHILE** $END = false$ and $UT_i = true$ **DO**
(2) Communicate with $\forall s_j \in S_i^u$ to obtain the strategy of s_j;
(3) Based on the definition of utility function, compute:
(4) $\Delta_i(\mathbf{a}) = \max_{a_i'} \left(u_i(a_i', \mathbf{a}_{-i}) - u_i(a_i, \mathbf{a}_{-i})\right)$
(5) and $br_i(\mathbf{a}) = \arg\max_{a_i'} \left(u_i(a_i', \mathbf{a}_{-i}) - u_i(a_i, \mathbf{a}_{-i})\right)$;
(6) **IF** $\Delta_i(\mathbf{a}) > 0$ **THEN**
(7) $a_i \leftarrow br_i(\mathbf{a})$; Sending $Update_i \leftarrow true$ to the system;
(8) **ELSE**
(9) Sending $Update_i \leftarrow false$ to the system;
(10) **END IF**
(11) **END WHILE**

Algorithm 2: Asynchronous distributed algorithm (ADA).

When Algorithm 2 proceeds, *END* is false. There is a sensor s_i updating its strategy and $\Delta_i(\mathbf{a}(t)) > 0$. Since direction scheduling game is a potential game, based on (1) and $\Phi(\mathbf{a}) = Z(\mathbf{a})$, $\Delta_i(\mathbf{a}(t)) > 0$ results in the increase of $Z(\mathbf{a})$. Thus, network coverage *monotonically increases* with the execution of Algorithm 2. Since the coverage objective function $Z(\mathbf{a})$ is *finite*, Algorithm 2 finally converges. At the time, no sensors are able to update strategies to increase utility and a pure Nash equilibrium is achieved. □

6. The Coverage Performance Analysis of Distributed Algorithms

A direction scheduling game is a potential game with potential function $\Phi(\mathbf{a}) = Z(\mathbf{a})$. Hence, network coverage strictly increases with execution of both synchronous and asynchronous distributed algorithms. Furthermore, both synchronous and asynchronous distributed algorithms finally converge to a stable profile, that is, a pure Nash equilibrium of direction scheduling game. Moreover, from the results of Theorem 16, the optimal solution of dKC-MCB is actually a pure Nash equilibrium of DSG. However, pure Nash equilibria are not unique for a DSG. Both synchronous and asynchronous distributed algorithms are not guaranteed to converge to optimal coverage. Thus, we should obtain explicit bounds on the coverage performance of both synchronous and asynchronous distributed algorithms. In terms of algorithmic game theory, we obtain the price of anarchy (PoA) [17] of direction scheduling game to analyze the coverage performance of proposed distributed algorithms.

Definition 18. The price of anarchy of a direction scheduling game DSG is defined as

$$\text{PoA (DSG)} = \max_{\mathbf{a}^* \in \mathbf{A}} \frac{Z\left(\mathbf{a}^{\text{OPT}}\right)}{Z\left(\mathbf{a}^*\right)}, \tag{4}$$

where $Z(\mathbf{a}) = \sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|$, that is, the objective function. $\mathbf{a}^{\text{OPT}}$ is the optimal solution of $Z(\cdot)$. $\mathbf{a}^*$ is a pure Nash equilibrium of a DSG.

Intuitively, a PoA is the ratio of optimal coverage and the worst possible Nash equilibrium coverage. Theorem 19 presents the explicit bounds on PoA of a DSG strongly depending on the submodularity [19] of coverage utility function of a DSG.

Theorem 19. *The upper bound of price of anarchy of a DSG is 2.*

We give the proof of Theorem 19 in Appendix B.

From the result of Theorem 19, we know that at least 1/2 optimal coverage can be obtained when synchronous and asynchronous distributed algorithms terminate.

7. Simulation Results

In this section, we evaluate the coverage performances and convergence of SDA and ADA algorithms through simulations. There are two measures for the coverage performance of algorithms. The first is the average coverage rate (ACR) of coverage optimization algorithms. ACR is an average of the coverage rate (CR) of each time slot which is the ratio between the number of targets covered by sensor networks and the number of targets located in a target area. The second is the coverage stability (CS) of coverage optimization algorithms which is measured by the variance of CRs of all time slots. The convergence is evaluated by the speed of convergence of SDA and ADA algorithms to a pure Nash equilibrium.

7.1. Experimental Demonstration of the Coverage Optimization. As an intuitive demonstration of distributed algorithms, Figure 1 shows the snapshots of coverage results of a random deployment and a SDA algorithm. In order to make the results accessible to readers, the small-scale simulations parameters are used in the demonstration. In this demonstration, $K = 3$ and each sensor has 7 sensing directions and can allocate at most 2 working directions into 3 time slots. The subfigures (a), (b), and (c) of Figure 1 show coverage results of random deployment in 3 different time slots. The subfigures (d), (e), and (f) of Figure 1 show

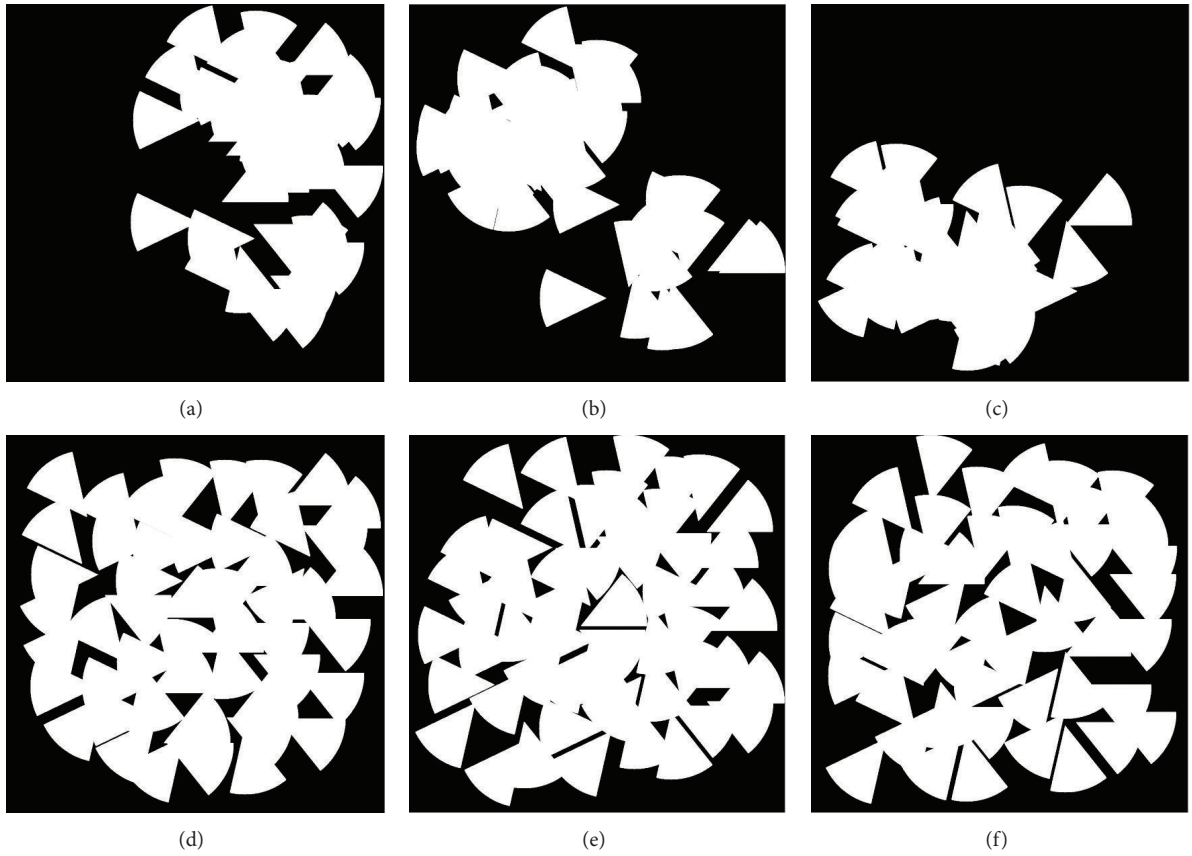


Figure 1: The snapshots of coverage results of the random deployment and the SDA algorithm.

coverage results in 3 different time slots when SDA algorithm converges to a pure Nash equilibrium. As we can see from the results shown in Figure 1, coverage results of SDA algorithm obviously outperform results of random deployment.

7.2. Coverage Performance of the Distributed Algorithms

7.2.1. Average Coverage Rate of the Distributed Algorithms. In order to evaluate the effectiveness of algorithms, we compare the coverage performance of SDA and ADA algorithms with the RANDOM algorithm and MCBLC-G algorithm. In a RANDOM algorithm, each sensor allocates randomly working directions into time slots under the constraint of its lifetime. MCBLC-G is a centralized greedy algorithm for the problem of minimum coverage breach which is developed by Yang et al. in [11]. The experiments are set up by the following settings. The target area is a 10×10 size area where 50 targets are randomly located. We evaluate the coverage performance of the SDA, ADA, RANDOM, and MCBLC-G algorithms by randomly deployed $n = 100, 125, 150, 175, 200, 225, 250, 275, 300$ sensors with $R_s = 2$ in target area, respectively. Total lifetime TL is divided equally into 6 time slots; that is, $K = 6$. Each sensor at most chooses 3 time slots to allocate its working directions; that is, $l_i \leq 3$. In order to guarantee the reliability of simulations, for each value of n, we repeat the simulations by 50 times. The ACR is finally computed by $((\sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|)/K)/50$.

Figure 2 shows coverage performances of the SDA, ADA, RANDOM, and MCBLC-G algorithms. As we can see from Figure 2, SDA, ADA, and MCBLC-G algorithms provide similar coverage performances which are obviously better than results of the RANDOM algorithm. However, MCBLC-G is a centralized algorithm and the SDA and ADA algorithms are distributed algorithms.

Since the potential function of direction scheduling game coincides with the coverage objective function of the problem of dKC-MCB, the coverage of sensor network increases along with the each sensor's utility increasing when SDA and ADA algorithms proceed. We can see the results from Figures 3 and 4.

7.2.2. Coverage Stability of the Distributed Algorithms. The coverage stability is another important performance measure of coverage optimization algorithms. A coverage optimization algorithm with good coverage stability can guarantee well-balanced coverage performance for every time slot. Figure 5 shows the coverage rate variance of SDA, ADA, RANDOM, and MCBLC-G algorithms. As we can see from the results of Figure 5, RANDOM algorithm randomly allocates working directions into time slots. Coverage rates among time slots cannot be guaranteed well balanced.

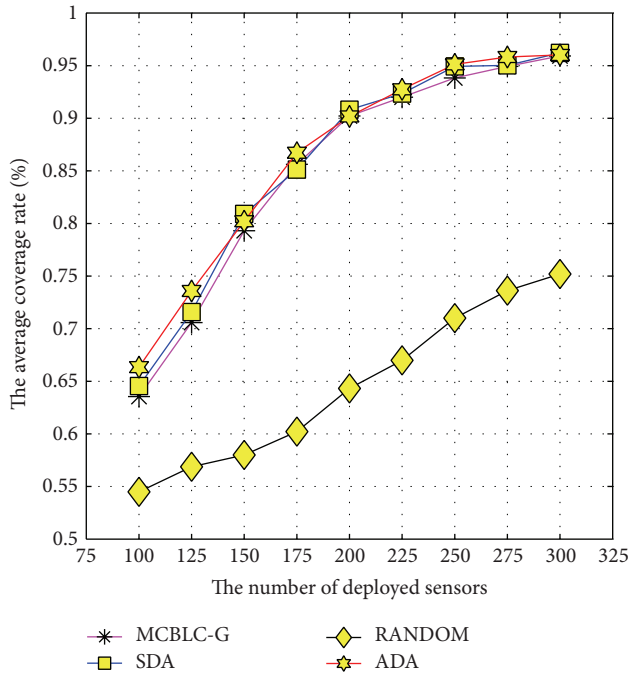


Figure 2: The comparison of coverage performance of SDA, ADA, RANDOM, and MCBLC-G algorithms.

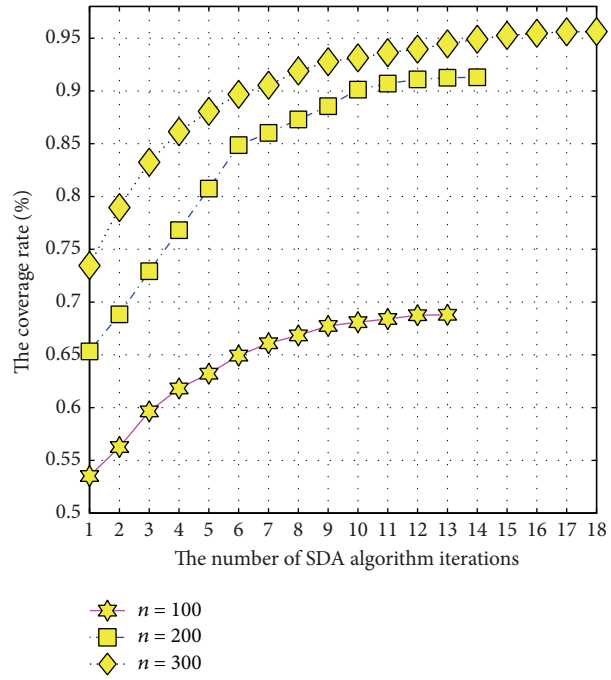


Figure 3: The increasing of coverage performance of SDA algorithms.

This results in a high coverage rate variance for RANDOM algorithm. MCBLC-G algorithm has a lower coverage rate variance than RANDOM. However, the coverage rate variance of MCBLC-G is influenced by the sequence of greedy selection of sensors. Compared with RANDOM and MCBLC-G algorithms, SDA and ADA provide well-balanced coverage performance.

7.3. Convergence of the Distributed Algorithms. A SDA (ADA) algorithm terminates when the algorithm converges to a pure Nash equilibrium. The speed of convergence to a pure Nash equilibrium of a SDA (ADA) algorithm mainly depends on two factors: the first is the number of sensors deployed in a target area. The second is the number of time slots K. As shown in Figure 6, the number of iterations for SDA algorithm increases with the number of sensors deployed in a target area. At the same time, given the number of deployed sensors, the number of iterations for SDA algorithm increases with the number of time slots K. As the number of K increases, each sensor has more choices when it decides to allocate working directions. This leads to more complicated interactions among sensors and prolongs the convergence dynamics. Similar convergence results of ADA algorithm are shown in Figure 7.

8. Conclusions

Direction scheduling algorithms with energy efficiency are always important for directional sensor network. Since directional sensors are energy constrained, distributed direction scheduling algorithms need to be exploited where a sensor takes direction scheduling decisions independently, based purely on communications with its neighbors. In this paper, the problem of directional K-Cover for minimum coverage breach (dKC-MCB) is formulated as a game: direction scheduling game (DSG). Both synchronous and asynchronous game-theoretic based distributed algorithms are proposed for solving dKC-MCB, respectively. The coverage performance of distributed algorithms is analyzed from a theoretic perspective and the explicit bounds on the coverage performance are presented. Experimental results show that our proposed algorithms can provide a near-optimal and well-balanced solution to the problem of dKC-MCB.

Game theory, particularly potential game, is beginning to emerge as a powerful tool for the design and analysis of distributed optimization algorithm [13]. However, many research challenges still remain unanswered, especially for the development of a systematic methodology for the design of distributed optimization functions that satisfies virtually any degree of locality while ensuring the desirability of the resulting equilibria. As future research work, we plan to investigate a systematic approach to distributed optimization using the framework of potential games and apply this approach to solve various real application problems.

Appendices

A. Proof of Theorems 12, 13, and 14

A.1. Proof of Theorem 12. Direction scheduling game is a potential game with the potential function $Z(\mathbf{a}) = \sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|$.

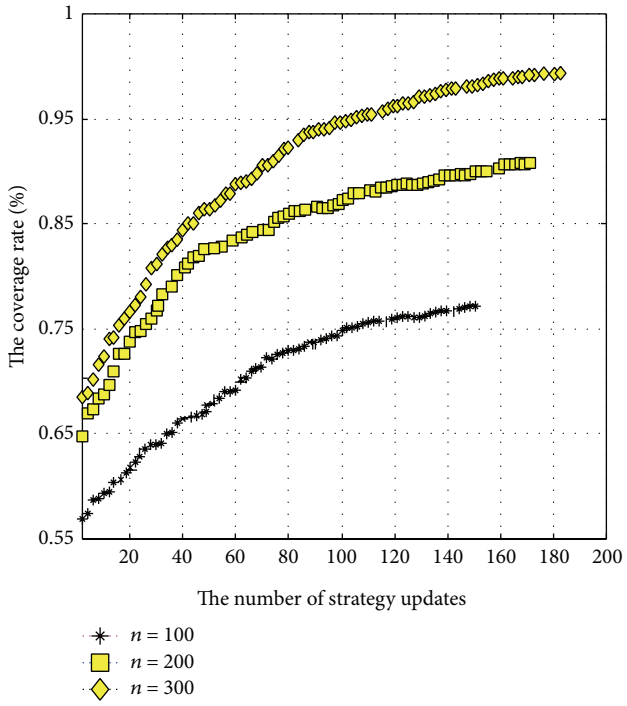


Figure 4: The increasing of coverage performance of ADA algorithms.

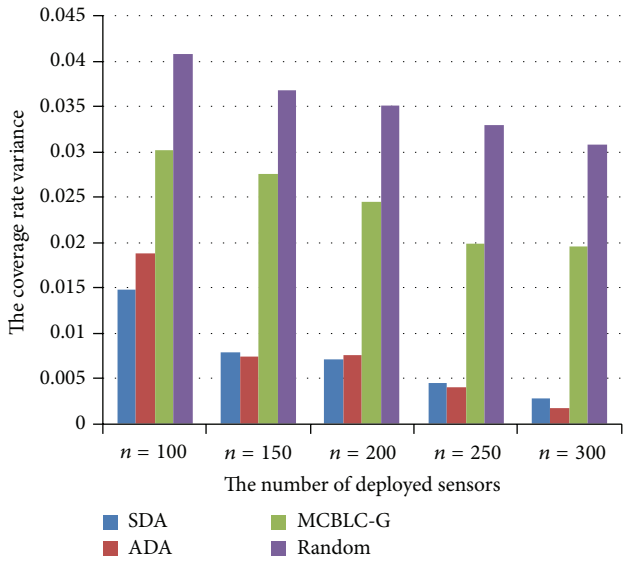


Figure 5: The comparison of coverage stability of SDA, ADA, RANDOM, and MCBLC-G algorithms.

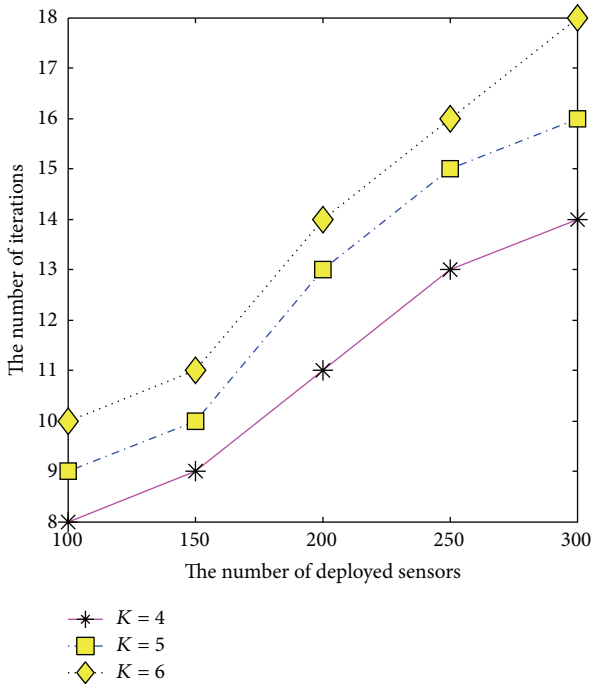


Figure 6: Convergence speed of SDA algorithms.

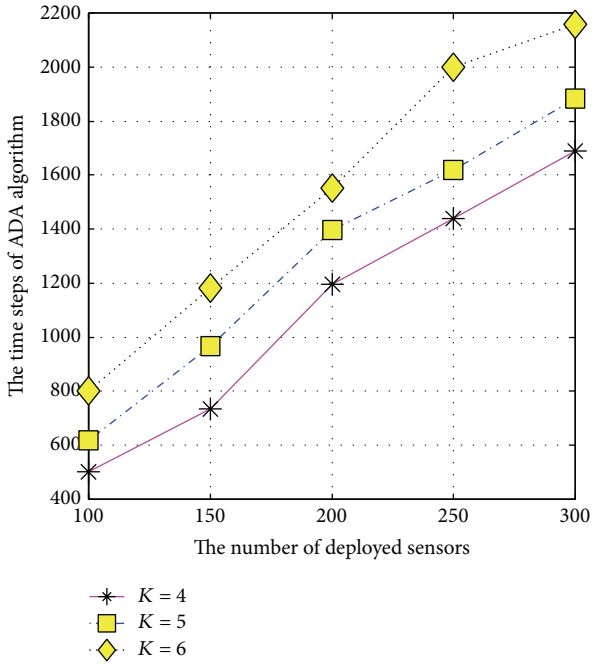


Figure 7: Convergence speed of ADA algorithms.

Proof. Given a game G, to prove G is a class of potential games, a potential function $\Phi(\cdot)$ should be constructed, such that for any tow profiles $\mathbf{a} = (a_i, \mathbf{a}_{-i})$ and $\mathbf{a}' = (a_i', \mathbf{a}_{-i})$ we always have

$$u_i\left(a_i', \mathbf{a}_{-i}\right) - u_i\left(a_i, \mathbf{a}_{-i}\right) = \Phi\left(a_i', \mathbf{a}_{-i}\right) - \Phi\left(a_i, \mathbf{a}_{-i}\right). \quad \text{(A.1)}$$

We prove that, for a direction scheduling game with utility function defined by (3), the potential function is $\Phi(\mathbf{a}) = Z(\mathbf{a})$.

Note that $\mathbf{a}' = (a_i', \mathbf{a}_{-i})$ and $\mathbf{a} = (a_i, \mathbf{a}_{-i})$ are only different from the strategy of sensor s_i since s_i at most activates a direction in a time slot. Thus, in a time slot t_k, the working directions $A_k^{\mathbf{a}'}$ differ from $A_k^{\mathbf{a}}$ in the following four possible cases.

(a) A sensor s_i schedules its working direction from $d_{i,p}$ at a profile $\mathbf{a}$ to $d_{i,q}$ at a profile $\mathbf{a}'$. In this case, $A_k^{\mathbf{a}'} = \{\{A_k^{\mathbf{a}} \setminus \{d_{i,p}\}\} \cup \{d_{i,q}\}\}$.

(b) s_i schedules its working direction from $d_{i,p}$ at a profile $\mathbf{a}$ to $\emptyset$ at a profile $\mathbf{a}'$. Thus, $A_k^{\mathbf{a}'} = \{A_k^{\mathbf{a}} \setminus \{d_{i,p}\}\}$.

(c) s_i schedules its working direction from $\emptyset$ at a profile $\mathbf{a}$ to $d_{i,q}$ at a profile $\mathbf{a}'$. Thus, $A_k^{\mathbf{a}'} = \{A_k^{\mathbf{a}} \cup \{d_{i,q}\}\}$.

(d) s_i maintains an equal schedule between $\mathbf{a}$ and $\mathbf{a}'$. Thus, $A_k^{\mathbf{a}'} = A_k^{\mathbf{a}}$.

Without loss of generality, at $\mathbf{a}$ and $\mathbf{a}'$, we suppose that $A_1^{\mathbf{a}'} = \{\{A_1^{\mathbf{a}} \setminus \{d_{i,p}\}\} \cup \{d_{i,q}\}\}$, $A_2^{\mathbf{a}'} = \{A_2^{\mathbf{a}} \setminus \{d_{i,p}\}\}$, $A_3^{\mathbf{a}'} = \{A_3^{\mathbf{a}} \cup \{d_{i,q}\}\}$, and $A_k^{\mathbf{a}'} = A_k^{\mathbf{a}}, k = 3, 4, \ldots, K$. Therefore, the potential functions of $\mathbf{a}$ and $\mathbf{a}'$ are, respectively, shown as follows:

$$\begin{aligned}
\Phi(\mathbf{a}) &= \sum_{k=1}^{L} F\left(A_k^{\mathbf{a}}\right) \\
&= F\left(A_1^{\mathbf{a}}\right) + F\left(A_2^{\mathbf{a}}\right) + F\left(A_3^{\mathbf{a}}\right) + F\left(A_4^{\mathbf{a}}\right) + \cdots + F\left(A_L^{\mathbf{a}}\right) \\
\Phi\left(\mathbf{a}'\right) &= \sum_{k=1}^{L} F\left(A_k^{\mathbf{a}'}\right) \\
&= F\left(A_1^{\mathbf{a}'}\right) + F\left(A_2^{\mathbf{a}'}\right) + F\left(A_3^{\mathbf{a}'}\right) \\
&\quad + F\left(A_4^{\mathbf{a}'}\right) + \cdots + F\left(A_L^{\mathbf{a}'}\right) \\
&= F\left(A_1^{\mathbf{a}'} = \left\{\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\} \cup \left\{d_{i,q}\right\}\right\}\right) \\
&\quad + F\left(A_2^{\mathbf{a}'} = \left\{A_2^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right) \\
&\quad + F\left(A_3^{\mathbf{a}'} = \left\{A_3^{\mathbf{a}} \cup \left\{d_{i,q}\right\}\right\}\right) + F\left(A_4^{\mathbf{a}'} = A_4^{\mathbf{a}}\right) \\
&\quad + \cdots + F\left(A_L^{\mathbf{a}'} = A_L^{\mathbf{a}}\right).
\end{aligned} \tag{A.2}$$

Thus, based on (A.2) we have

$$\begin{aligned}
\Phi\left(\mathbf{a}'\right) - \Phi(\mathbf{a}) &= \sum_{k=1}^{L} F\left(A_k^{\mathbf{a}'}\right) - \sum_{k=1}^{L} F\left(A_k^{\mathbf{a}}\right) \\
&= \left[F\left(A_1^{\mathbf{a}'}\right) - F\left(A_1^{\mathbf{a}}\right)\right] + \left[F\left(A_2^{\mathbf{a}'}\right) - F\left(A_2^{\mathbf{a}}\right)\right] \\
&\quad + \left[F\left(A_3^{\mathbf{a}'}\right) - F\left(A_3^{\mathbf{a}}\right)\right] + \left[F\left(A_4^{\mathbf{a}'}\right) - F\left(A_4^{\mathbf{a}}\right)\right] \\
&\quad + \cdots + \left[F\left(A_L^{\mathbf{a}'}\right) - F\left(A_L^{\mathbf{a}}\right)\right] \\
&= \left[F\left(\left\{\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\} \cup \left\{d_{i,q}\right\}\right\}\right) - F\left(A_1^{\mathbf{a}}\right)\right] \\
&\quad + \left[F\left(\left\{A_2^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right) - F\left(A_2^{\mathbf{a}}\right)\right] \\
&\quad + \left[F\left(\left\{A_3^{\mathbf{a}} \cup \left\{d_{i,q}\right\}\right\}\right) - F\left(A_3^{\mathbf{a}}\right)\right] + 0.
\end{aligned} \tag{A.3}$$

In addition, based on the definition of utility function, we obtain the utility of s_i at $\mathbf{a}$,

$$\begin{aligned}
u_i(\mathbf{a}) &= \left[F\left(A_1^{\mathbf{a}}\right) - F\left(\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right)\right] \\
&\quad + \left[F\left(A_2^{\mathbf{a}}\right) - F\left(\left\{A_2^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right)\right],
\end{aligned} \tag{A.4}$$

and at $\mathbf{a}'$, respectively,

$$\begin{aligned}
u_i\left(\mathbf{a}'\right) &= \begin{bmatrix} F\left(\left\{\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\} \cup \left\{d_{i,q}\right\}\right\}\right) \\ -F\left(\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right) \end{bmatrix} \\
&\quad + \left[F\left(\left\{A_3^{\mathbf{a}} \cup \left\{d_{i,q}\right\}\right\}\right) - F\left(A_3^{\mathbf{a}}\right)\right].
\end{aligned} \tag{A.5}$$

Thus, based on the results of (A.4) and (A.5), the difference of utility of s_i between $\mathbf{a}$ and $\mathbf{a}'$ is as follows:

$$\begin{aligned}
u_i\left(\mathbf{a}'\right) - u_i(\mathbf{a}) &= \begin{bmatrix} F\left(\left\{\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\} \cup \left\{d_{i,q}\right\}\right\}\right) \\ -F\left(\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right) \end{bmatrix} \\
&\quad + \left[F\left(\left\{A_3^{\mathbf{a}} \cup \left\{d_{i,q}\right\}\right\}\right) - F\left(A_3^{\mathbf{a}}\right)\right] \\
&\quad - \left[F\left(A_1^{\mathbf{a}}\right) - F\left(\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right)\right] \\
&\quad - \left[F\left(A_2^{\mathbf{a}}\right) - F\left(\left\{A_2^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right)\right] \\
&= \left[F\left(\left\{\left\{A_1^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\} \cup \left\{d_{i,q}\right\}\right\}\right) - F\left(A_1^{\mathbf{a}}\right)\right] \\
&\quad + \left[F\left(\left\{A_2^{\mathbf{a}} \setminus \left\{d_{i,p}\right\}\right\}\right) - F\left(A_2^{\mathbf{a}}\right)\right] \\
&\quad + \left[F\left(\left\{A_3^{\mathbf{a}} \cup \left\{d_{i,q}\right\}\right\}\right) - F\left(A_3^{\mathbf{a}}\right)\right].
\end{aligned} \tag{A.6}$$

From the results of (A.3) and (A.6), we obtain $u_i(a_i', \mathbf{a}_{-i}) - u_i(a_i, \mathbf{a}_{-i}) = \Phi(a_i', \mathbf{a}_{-i}) - \Phi(a_i, \mathbf{a}_{-i})$. Therefore, we prove that a direction scheduling game is a potential game with potential function $\Phi(\mathbf{a}) = Z(\mathbf{a})$. □

A.2. Proof of Theorem 13. A pure Nash equilibrium of direction scheduling game is a local optimal solution of the problem of Set K-Cover for minimum coverage breach.

Proof. Let $\mathbf{a}^*$ be a pure Nash equilibrium profile of DSG. By the definition of Nash equilibrium, at the profile $\mathbf{a}^*$, there is no sensor that can unilaterally deviate from the profile $\mathbf{a}^*$ by changing its scheduling strategy to increase its individual coverage utility. Sine DSG is a potential game of which the potential function $\Phi(\cdot)$ is consistent with $Z(\cdot)$, that is, the optimization objective function of dKC-MCB, in a local area around the profile $\mathbf{a}^*$, there is no profile $\mathbf{a}'$ such that $Z(\mathbf{a}^*) < Z(\mathbf{a}')$. Thereby, the Nash equilibrium profile $\mathbf{a}^*$ is local optimal of the optimization objective function of the problem of dKC-MCB. □

A.3. Proof of Theorem 14. An optimal solution to the problem of dKC-MCB is a pure Nash equilibrium.

Proof. Let $\mathbf{a}^{\mathrm{OPT}} = \arg\max_{\mathbf{a}} Z(\mathbf{a}) = \arg\max_{\mathbf{a}} \sum_{k=1}^{K} t_k |F(A_k^{\mathbf{a}})|$ be the optimal solution profile of dKC-MCB. At the profile $\mathbf{a}^{\mathrm{OPT}}$, a sensor s_i unilaterally deviates from the profile $\mathbf{a}^{\mathrm{OPT}}$ by changing its scheduling strategy to a profile $\mathbf{a}'$. Since $\mathbf{a}^{\mathrm{OPT}}$ is the optimal solution profile, we obtain $Z(\mathbf{a}^{\mathrm{OPT}}) \geq Z(\mathbf{a}')$. Moreover, since a DSG is a potential game of which the potential function is consistent with $Z(\cdot)$, $Z(\mathbf{a}^{\mathrm{OPT}}) \geq Z(\mathbf{a}')$ implies that, at the profile $\mathbf{a}^{\mathrm{OPT}}$, there is no sensor that can unilaterally deviate from the profile $\mathbf{a}^{\mathrm{OPT}}$ to increase its utility. By the definition of Nash equilibrium, we know the optimal solution profile $\mathbf{a}^{\mathrm{OPT}}$ is a pure Nash equilibrium of a DSG. □

B. Proof of Theorem 19

The upper bound of price of anarchy of a direction scheduling game is 2.

Proof. Let $\mathbf{a}^* = (a_1^*, \ldots, a_i^*, \ldots, a_n^*)$ be a pure Nash equilibrium of a DSG and let $\mathbf{a}^{\mathrm{OPT}} = (a_1^{\mathrm{OPT}}, \ldots, a_i^{\mathrm{OPT}}, \ldots, a_n^{\mathrm{OPT}})$ be the optimal profile of the problem of Set K-Cover for minimum coverage breach. Let $(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}})$ denote a compound profile where each sensor s_i can take both a_i^* and a_i^{OPT} strategies. At a compound profile $(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}})$, the network coverage is $Z(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}}) = \sum_{k=1}^{K} F(A_k^{(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}})})$. Since the network coverage function $F(\cdot)$ is a monotone function on the set of directions, we have $Z(\mathbf{a}^*) \leq Z(\mathbf{a}^{\mathrm{OPT}}) \leq Z(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}})$.

Denote $\Delta Z_i(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}})$ by

$$\begin{aligned}
&\Delta Z_i\left(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}}\right) \\
&\quad = Z\left(\mathbf{a}^* \oplus \left(a_1^{\mathrm{OPT}}, \ldots, a_{i-1}^{\mathrm{OPT}}, a_i^{\mathrm{OPT}}, a_{i+1}^*, \ldots, a_n^*\right)\right) \\
&\qquad - Z\left(\mathbf{a}^* \oplus \left(a_1^{\mathrm{OPT}}, \ldots, a_{i-1}^{\mathrm{OPT}}, a_i^*, a_{i+1}^*, \ldots, a_n^*\right)\right) \\
&\quad = Z\left(a_1^{\mathrm{OPT}} \oplus a_1^*, \ldots, a_{i-1}^{\mathrm{OPT}} \oplus a_{i-1}^*, a_i^{\mathrm{OPT}} \oplus a_i^*, a_{i+1}^*, \ldots, a_n^*\right) \\
&\qquad - Z\left(a_1^{\mathrm{OPT}} \oplus a_1^*, \ldots, a_{i-1}^{\mathrm{OPT}} \oplus a_{i-1}^*, a_i^*, a_{i+1}^*, \ldots, a_n^*\right).
\end{aligned} \tag{B.1}$$

Without loss of generality, assume that, at the profile $\mathbf{a}^{\mathrm{OPT}}$, s_i allocates directions in time slots indexed by $(p_1, p_2, \ldots, p_s)$. Thus, s_i's strategy a_i^{OPT} is denoted as follows:

$$\begin{aligned}
a_i^{\mathrm{OPT}} = \Big\{ &\left\langle t_{p_1}, d_{i,q_1}^{\mathbf{a}^{\mathrm{OPT}}} \right\rangle, \ldots, \left\langle t_{p_s}, d_{i,q_s}^{\mathbf{a}^{\mathrm{OPT}}} \right\rangle \mid t_{p_k} \in T, \\
&d_{i,q_k}^{\mathbf{a}^{\mathrm{OPT}}} \in D_i; \ \sum t_{p_k} \leq l_i \Big\}.
\end{aligned} \tag{B.2}$$

Let $\mathbf{a}' = (a_1^*, \ldots, a_{i-1}^*, a_i^{\mathrm{OPT}}, a_{i+1}^*, \ldots, a_n^*)$; that is, s_i's strategy a_i^* is replaced by a_i^{OPT} in a pure Nash equilibrium $\mathbf{a}^*$. Let $\hat{\mathbf{a}} = (a_1^*, \ldots, a_{i-1}^*, a_i^* \oplus a_i^{\mathrm{OPT}}, a_{i+1}^*, \ldots, a_n^*)$; that is, s_i's strategy a_i^* is replaced by a compound strategy $a_i^* \oplus a_i^{\mathrm{OPT}}$ in a pure Nash equilibrium $\mathbf{a}^*$. Thus, for for all t_{p_k}, $p_k \in (p_1, p_2, \ldots, p_s)$ and direction $d_{i,q_k}^{\mathbf{a}^{\mathrm{OPT}}} \in D_i$, we have $(A_{p_k}^{\mathbf{a}'} \setminus \{d_{i,q_k}^{\mathbf{a}^{\mathrm{OPT}}}\}) \subseteq A_{p_k}^{\mathbf{a}^*} \subseteq A_{p_k}^{\hat{\mathbf{a}}}$. Since the network coverage function is submodular function [19] on the set of directions,

$$\begin{aligned}
&F\left(A_{p_k}^{\hat{\mathbf{a}}}\right) - F\left(A_{p_k}^{\hat{\mathbf{a}}} \setminus \left\{d_{i,q_k}^{\mathbf{a}^{\mathrm{OPT}}}\right\}\right) \\
&\quad \leq F\left(A_{p_k}^{\mathbf{a}'}\right) - F\left(A_{p_k}^{\mathbf{a}'} \setminus \left\{d_{i,q_k}^{\mathbf{a}^{\mathrm{OPT}}}\right\}\right).
\end{aligned} \tag{B.3}$$

Based on the results of formula (B.3), we know that

$$\begin{aligned}
&\Delta Z_i\left(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}}\right) \\
&\quad = Z\left(\mathbf{a}^* \oplus \left(a_1^{\mathrm{OPT}}, \ldots, a_{i-1}^{\mathrm{OPT}}, a_i^{\mathrm{OPT}}, a_{i+1}^*, \ldots, a_n^*\right)\right) \\
&\qquad - Z\left(\mathbf{a}^* \oplus \left(a_1^{\mathrm{OPT}}, \ldots, a_{i-1}^{\mathrm{OPT}}, a_i^*, a_{i+1}^*, \ldots, a_n^*\right)\right) \\
&\quad = Z\left(a_1^{\mathrm{OPT}} \oplus a_1, \ldots, a_{i-1}^{\mathrm{OPT}} \oplus a_{i-1}, a_i^{\mathrm{OPT}} \oplus a_i, a_{i+1}^*, \ldots, a_n^*\right) \\
&\qquad - Z\left(a_1^{\mathrm{OPT}} \oplus a_1, \ldots, a_{i-1}^{\mathrm{OPT}} \oplus a_{i-1}, a_i^*, a_{i+1}^*, \ldots, a_n^*\right) \\
&\quad \leq u_i\left(a_i^{\mathrm{OPT}}, \mathbf{a}_{-i}^*\right).
\end{aligned} \tag{B.4}$$

Since $\mathbf{a}^*$ is a pure Nash equilibrium, suppose that other sensors' strategies are maintained to be unchanged, and any the coverage utility of sensor s_i is decreased by replacing a_i^* with a_i^{OPT}. Thus, $u_i(a_i^{\mathrm{OPT}}, \mathbf{a}_{-i}^*) \leq u_i(a_i^*, \mathbf{a}_{-i}^*)$. Therefore, we further obtain

$$\begin{aligned}
&Z\left(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}}\right) - Z\left(\mathbf{a}^*\right) \\
&\quad = \sum_{i=1}^{n} \Delta Z_i\left(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}}\right) \leq \sum_{i=1}^{n} u_i\left(\mathbf{a}^*\right).
\end{aligned} \tag{B.5}$$

Since $Z(\mathbf{a}^{\mathrm{OPT}}) \leq Z(\mathbf{a}^* \oplus \mathbf{a}^{\mathrm{OPT}})$, we have

$$\begin{aligned}
&Z\left(\mathbf{a}^{\mathrm{OPT}}\right) - Z\left(\mathbf{a}^*\right) \leq \sum_{i=1}^{n} u_i\left(\mathbf{a}^*\right) \\
&\quad \Longrightarrow \frac{Z\left(\mathbf{a}^{\mathrm{OPT}}\right)}{Z\left(\mathbf{a}^*\right)} \leq 1 + \frac{\sum_{i=1}^{n} u_i\left(\mathbf{a}^*\right)}{Z\left(\mathbf{a}^*\right)}.
\end{aligned} \tag{B.6}$$

In a direction scheduling game, the utility function of a sensor is defined by the marginal contribution to network coverage; therefore, we have $\sum_{i=1}^{n} u_i(\mathbf{a}^*) \leq Z(\mathbf{a}^*)$. Thereby, we obtain

$$\begin{aligned}
\mathrm{PoA\ (DSG)} &= \max_{\mathbf{a}^*} \frac{Z\left(\mathbf{a}^{\mathrm{OPT}}\right)}{Z\left(\mathbf{a}^*\right)} \\
&\leq 1 + \frac{\sum_{i=1}^{n} u_i\left(\mathbf{a}^*\right)}{Z\left(\mathbf{a}^*\right)} \leq 2.
\end{aligned} \tag{B.7}$$

□

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61163003), the Yunnan Provincial Foundation for Leaders of Disciplines in Science and Technology (2012HB004), the Natural Science Foundation of Yunnan Province (2011FB020), the Foundation of Key Program of Department of Education of Yunnan Province (2013Z049), the Foundation of Development Program for Backbone Teachers of Yunnan University (XT412003), and the Foundation of the Key Laboratory of Software Engineering of Yunnan Province (2012SE303, 2012SE205).

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[2] B. Wang, "Coverage problems in sensor networks: a survey," *ACM Computing Surveys*, vol. 43, no. 4, article 32, 2011.

[3] D. Tao and H. D. Ma, "Coverage control algorithms for directional sensor networks," *Journal of Software*, vol. 22, no. 10, pp. 2317–2334, 2011.

[4] S. Slijepcevic and M. Potkonjak, "Power efficient organization of wireless sensor networks," in *Proceedings of the IEEE International Conference on Communications (ICC '01)*, pp. 472–476, IEEE Computer Society Press, Helsinki, Finland, June 2000.

[5] J. Ai and A. A. Abouzeid, "Coverage by directional sensors in randomly deployed wireless sensor networks," *Journal of Combinatorial Optimization*, vol. 11, no. 1, pp. 21–41, 2006.

[6] Y. Cai, W. Lou, M. Li, and X.-Y. Li, "Energy efficient target-oriented scheduling in directional sensor networks," *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1259–1274, 2009.

[7] J. Wen, L. Fang, J. Jiang, and W. H. Dou, "Coverage optimizing and node scheduling in directional wireless sensor networks," in *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pp. 1–4, Dalian, China, October 2008.

[8] M. X. Cheng, L. Ruan, and W. Wu, "Coverage breach problems in bandwidth-constrained sensor networks," *ACM Transactions on Sensor Networks*, vol. 3, no. 2, article 12, 2007.

[9] Z. Abrams, A. Goel, and S. A. Plotkin, "Set K-cover algorithms for energy efficient monitoring in wireless sensor networks," in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks (IPSN '04)*, pp. 424–432, ACM Press, Berkeley, Calif, USA, April 2004.

[10] A. Deshpande, S. Khuller, A. Malekian, and M. Toossi, "Energy efficient monitoring in sensor networks," *Algorithmica*, vol. 59, no. 1, pp. 94–114, 2011.

[11] H.-Q. Yang, D.-Y. Li, and Z. Li, "Minimum coverage breach and maximum network lifetime in directional sensor networks," *Acta Electronica Sinica*, vol. 38, no. 2, pp. 138–142, 2010.

[12] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, Cambridge, Mass, USA, 1991.

[13] B. Yang and M. Johansson, "Distributed optimization and games: a tutorial overview," in *Networked Control Systems*, vol. 406 of *Lecture Notes in Control and Information Sciences*, pp. 109–148, Springer, London, UK, 2010.

[14] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[15] J. R. Marden, G. Arslan, and J. S. Shamma, "Cooperative control and potential games," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, vol. 39, no. 6, pp. 1393–1407, 2009.

[16] U. O. Candogan, I. Menache, A. Ozdaglar, and P. A. Parrilo, "Near-optimal power control in wireless networks: a potential game approach," in *Proceedings of the 29th Conference on Information Communications (INFOCOM '10)*, pp. 1954–1962, IEEE Computer Society Press, San Diego, Calif, USA, March 2010.

[17] E. Koutsoupias and C. H. Papadimitriou, "Worst-case equilibria," in *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science (STACS '99)*, pp. 404–413, 1999.

[18] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: efficient averaging for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1205–1216, 2008.

[19] S. Fujishige, *Submodular Functions and Optimization*, Elsevier, Amsterdam, The Netherlands, 2nd edition, 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 576401, 11 pages
http://dx.doi.org/10.1155/2014/576401

Research Article

Altruistic Backoff: Collision Avoidance for Receiver-Initiated MAC Protocols for Wireless Sensor Networks

Xenofon Fafoutis, Charalampos Orfanidis, and Nicola Dragoni

Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

Correspondence should be addressed to Xenofon Fafoutis; xefa@dtu.dk

Received 28 October 2013; Revised 20 January 2014; Accepted 7 May 2014; Published 21 May 2014

Academic Editor: Christos Verikoukis

In receiver-initiated medium access control (MAC) protocols for wireless sensor networks, communication is initiated by the receiver node which transmits beacons indicating its availability to receive data. In the case of multiple senders having traffic for a given receiver, such beacons form points where collisions are likely to happen. In this paper, we present *altruistic backoff* (AB), a novel collision avoidance mechanism that aims to avoid collisions before the transmission of a beacon. As a result of an early backoff, senders spend less time in idle listening waiting for a beacon, thus saving significant amounts of energy. We present an implementation of AB for Texas Instruments' eZ430-rf2500 sensor nodes and we evaluate its performance with simulations and experiments.

1. Introduction

Wireless sensor networks (WSNs) consist of multiple embedded networked wireless devices that are characterized by resource and power constraints. The medium access control (MAC) protocol is responsible for the establishment of a communication link between wireless devices. Its primary role is to coordinate access to and transmission over a medium common to several nodes. Furthermore, it plays a key role in the design of energy-efficient WSNs, as it controls the active and sleeping states of a node, known as *duty cycling*. The energy consumption of a wireless sensor node is dominated by the power needs of its radio component [1]. As a result, duty cycling the radio plays a fundamental role towards the realization of energy-efficient wireless sensor networks.

For a communication link to be established, both the receiver and the sender need to be simultaneously in an active state. Here, an important distinction needs to be made. In the case of single-hop star topologies and assuming that the receiver has sufficient energy resources to be continuously in active mode, establishing the link does not constitute a particular challenge. A duty-cycling sender will always find the receiver available to receive traffic. Related work, in this scenario, primarily builds upon the IEEE 802.15.4 standard [2], such as DQ-MAC [3].

In multihop topologies, on the other hand, both the sender and the receiver are duty cycling. This poses a particular problem of finding a rendezvous point between a sender and receiver in which both of the nodes are in an active state and a communication link can be established. In the literature there are three fundamental approaches to address this issue. In protocols that follow a *synchronous* approach, like S-MAC [4], T-MAC [5], and DSMAC [6], to mention only a few, nodes organize the active and sleeping states to overlap with each other. The beacon-enable mode of IEEE 802.15.4 and its extensions (e.g., NCCARQ-WSN [7]) can be also classified as synchronous MAC protocols.

Asynchronous schemes do not require synchronization, as the nodes sleep and wake up independently of the others. This leads to the need of techniques on deciding a rendezvous point for nodes to communicate. There are two fundamental asynchronous techniques, namely, the *sender-initiated* and the *receiver-initiated*. The basic technique used in a *sender-initiated asynchronous MAC scheme* is called preamble sampling, where the sender transmits a preamble to indicate that there is a pending need for communication. The receiver wakes up occasionally into the active state to

listen to such a preamble transmission. Once the preamble is detected, the receiver replies with a positive acknowledgment to the sender when the preamble transmission stops. This establishes a communication link between the sender and receiver. Most notable examples of MAC protocols that are based on the sender-initiated paradigm are WiseMAC [8], B-MAC [9], and X-MAC [10]. A thorough survey of sender-initiated schemes can be found in [11].

This paper focuses on the latter asynchronous approach, the *receiver-initiated* approach. In *receiver-initiated asynchronous MAC protocols*, the sender listens to the channel waiting for small beacons transmitted by the receiver. The receiver transmits the beacons, which are used by the sender to synchronize with the receiver, in accordance to its duty cycle. The receiver-initiated paradigm was originally introduced by Lin et al. in 2004 (RICER [12]) and became popular with RI-MAC [13] in 2008.

Contribution and Outline of the Paper. The key idea behind the receiver-initiated paradigm is that beacons constitute indirect transmission timeslots. Therefore, when multiple nodes contend for the same beacon, a collision is inevitable. Unless there are specific conditions so that receivers can provide the network with much more beacons than the generated data packets, receiver-initiated protocols are particularly vulnerable to collisions. In this paper, we present *altruistic backoff* (AB), a novel energy-efficient collision avoidance mechanism for receiver-initiated MAC protocols. AB manages to decrease the energy wasted in idle listening by detecting and resolving inevitable collisions before the beacon transmission. Additionally, we implemented the protocol on Texas Instruments' eZ430-rf2500 sensor nodes [14] and we evaluate its performance by comparing it to *random backoff* through simulations and experiments.

The remainder of the paper is organized as follows. Section 2 presents the receiver-initiated paradigm of communication along with previous work on collision avoidance. Section 3 summarizes the proposed protocol, AB. Section 4 evaluates the protocol using simulations. Section 5 provides the implementation details and Section 6 presents the experimental results. Lastly, Section 7 concludes the paper.

2. Collision Avoidance in Receiver-Initiated MAC Protocols

In this section, we present the receiver-initiated paradigm of communication between duty-cycling nodes. Furthermore, we briefly survey how existing MAC protocols, that incorporate the receiver-initiated paradigm, address the challenge of collision avoidance. Lastly, we motivate the necessity of our proposed solution by contrasting it with the commonly used approach in terms of energy efficiency.

2.1. The Receiver-Initiated Paradigm. Receiver-initiated MAC protocols use beacons to establish a link between duty-cycling nodes, as sketched in Figure 1. In particular, a node is usually in a sleeping state, in which its radio is turned off. Occasionally, it interrupts its sleep to transmit a small frame,

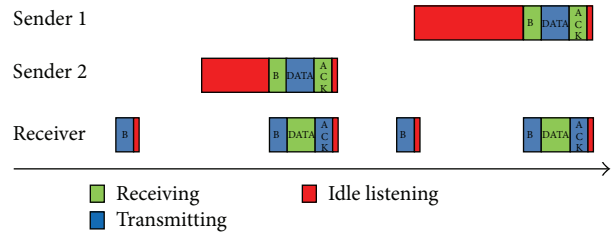


Figure 1: Receiver-initiated paradigm of communication. The sender is passively listening to the channel for a beacon (B) that initiates the communication. *Idle listening* indicates a source of energy consumption where the node is active, receiving, while the other side of the link is in sleeping mode.

called *beacon* (B), which indicates its availability to receive data. After the beacon transmission and for a predefined time, the node waits (with the radio tuned on) for a reply. In case of no reply, the node goes back to the sleeping state. A node with data to transmit interrupts its sleep and passively listens to the channel for a beacon that originates from the intended receiver. Upon reception of a beacon, data transmission follows, typically acknowledged by an additional control frame (ACK). The latter concludes the communication cycle and both nodes go to the sleeping state.

Since the publication of RI-MAC [13], several MAC protocols that build on the receiver-initiated paradigm have been proposed. Such protocols mostly focus on optimizing the performance of the network and/or extending some features, such as mitigating the cost of beaconing (e.g., A-MAC [15]), mitigating the time a node awaits for a beacon (e.g., EE-RI-MAC [16] and PW-MAC [17]), dynamically adapting the duty cycles (e.g., ODMAC [18] and CyMAC [19]), adding broadcasting support (ADB [20] and YA-MAC [21]), and adding multichannel support (DCM [22] and EM-MAC [23]). Despite their differences, all these MAC protocols are based on the same receiver-initiated communication paradigm.

A comparison between the receiver-initiated paradigm and other communication paradigms for duty-cycled nodes is out of the scope of paper. For such comparison, we refer the reader to related works [12, 13, 24].

In the literature, there is a line of MAC protocols that extend the paradigm with techniques to predict the wake-up event of the receiver, such as WideMAC [25] and PW-MAC [17]. Moreover, there are several other extensions to the paradigm that are incompatible with such prediction techniques. For instance, in ODMAC [18], the wake-up events are unpredictable as they are scheduled based on the available energy that can be harvested from the environment. The main disadvantages of the wake-up prediction techniques are related to the requirement to deal with clock drifts in the microcontrollers of the sensor nodes and to the fact that they hinder the ability of sensor nodes to autonomously and individually adapt their duty cycles to various environmental parameters. In such dynamic conditions, the wake-up event of the receiver is continuously changing with respect to the available energy or the network conditions. Furthermore,

MAC protocols typically follow randomization techniques that aim to avoid unwanted synchronizations (for instance, two nodes continuously transmit their beacon simultaneously) [13]. In this paper, we consider the generic version of the receiver-initiated paradigm, in which the wake-up events of the receiver are unknown to the sender.

2.2. Collision Avoidance. Collision avoidance in wireless networks was introduced because collision detection mechanisms, traditionally used in wired networks, are impossible. Detecting a collision while it is happening is not possible in wireless networks, because the radio is not able to transmit and receive simultaneously. Collided transmissions can only be detected by the receiver after their completion. Therefore, in high throughput wireless networks with large data packets, such as IEEE 802.11 [26], collisions lead to a significant throughput degradation.

The solution to this problem was given by avoiding collisions through *Random Backoff* (RB). The idea is that the protocol defines a time interval (*timeslot*) and a contention window (CW). Before transmitting, each node selects a random number, chosen uniformly between zero and CW − 1, and it delays the data transmission by that amount of timeslots while listening to the channel for other transmissions. If the channel remains idle, the data transmission follows. If the channel gets occupied by another transmission, the node freezes the timeslot counter and backs off. When the channel becomes idle again the node unfreezes the timeslot counter and the process is repeated until the counter reaches zero. At this point, the data transmission follows. As a result, unless two transmitters select the same random number, the collision is avoided. The size of CW is associated with a performance tradeoff. If its value is too small, the probability of two nodes selecting the same random number gets high. On the other hand, if its value is too high, the transmitters waste a lot of time in idle listening, leading to protocol overhead and throughput degradation. IEEE 802.11 distributed coordination function (DCF) [26] solves this problem by adapting CW to the level of contention. This mechanism works as follows. CW is initialized with a small value, which is doubled every time a collision occurs (with a maximum limit) and gets back to its minimum value after a successful transmission. This mechanism is called *binary exponential backoff* and results to a low CW in low contention that can quickly increase in the case of traffic bursts.

Receiver-initiated MAC protocols for WSNs inherited the principle of RB from traditional wireless protocols. RI-MAC [13] adopts a variation of BEB. The experiments conducted by the authors of RI-MAC have shown that due to the presence of the capture effect [27] in FM radios, also called cochannel interference tolerance, such a contending scenario does not necessarily lead to collisions. This property shows that the traditional assumption that a packet collision always results in data corruption is false. For this reason, senders in RI-MAC immediately transmit the data upon receiving a base beacon, without any backoff. The receiver listens for a short period of time after transmitting the beacon, known as the dwell time. Dwell time is determined by the current backoff window size. Concurrently, it measures the channel power level and processes the bit pattern received. If a valid data frame header is not detected in time and the measured power level indicates that a transmission is in progress, then this condition is classified as a collision. If a collision occurs, the receiver performs a clear channel assessment (CCA), waiting for the channel to be free. Once a clear channel is determined, the receiver transmits a beacon with a backoff window specified, informing the senders of the failed transmission. The senders that are waiting for an ACK use the backoff window specified in the beacon to perform a random backoff. The senders listen to the channel, while waiting for the random period to expire, before retransmitting the data. If a transmission from another sender is detected, the sender withholds the transmission and waits for an ACK beacon, before resuming with a new random backoff. If a collision happens again, the receiver increments the backoff window using the BEB strategy, until the maximum window size is reached. After that, both senders and receivers accept a failed transmission and go back to sleep, retrying at a later point in time.

In addition to RI-MAC, other receiver-initiated MAC protocols adopt variations of RB, including RICER [12], RC-MAC [28], YA-MAC [21], DCM [22], EM-MAC [23], and IRDT [29]. Intermittent receiver-driven data transmission (IRDT) [29] also incorporates two additional collision avoidance mechanisms. The first is based on the frequency of beacon transmissions. The idea is that by increasing the beacons, the senders are stochastically distributed into more beacons and the collision probability decreases. However, this solution can work only if the receivers are capable of offering their energy resources for forwarding more traffic. The other collision avoidance mechanism is based on data aggregation. By aggregating multiple data packets into larger frames, the total amount of attempted transmissions decreases; thus, the probability of a collision decreases. However, this approach has a negative impact on the delay of each individual data packet. The authors define two methods of collision avoidance with data aggregation: static and dynamic. According to the static method, the protocol uses a constant buffer of n packets. The node keeps collecting packets from other nodes and packets locally generated into the buffer. When the buffer is full, it is transmitted as a single MAC frame. According to the dynamic method, a sender with a single packet to transmit normally waits for a beacon. While waiting, it periodically transmits its own beacons in order to collect packets from neighbors. When the beacon is received, the sender transmits a single frame with as many packets as it managed to collect during that time.

Self-adapting RI-MAC (SARI-MAC) [30] introduces a collision avoidance mechanism through time slot reservation. After the beacon transmission, a contention window period follows during which the nodes pick a uniformly random slot to request for a timeslot reservation. At the end of the contention window, the receiver sends back to all the contending nodes a report with the reservations. Nodes transmit their data in the reserved timeslot, which is long enough for a data packet and the respective acknowledgment. Opportunistic cooperation MAC (OC-MAC) [31] indirectly decreases collisions by allowing senders to opportunistically

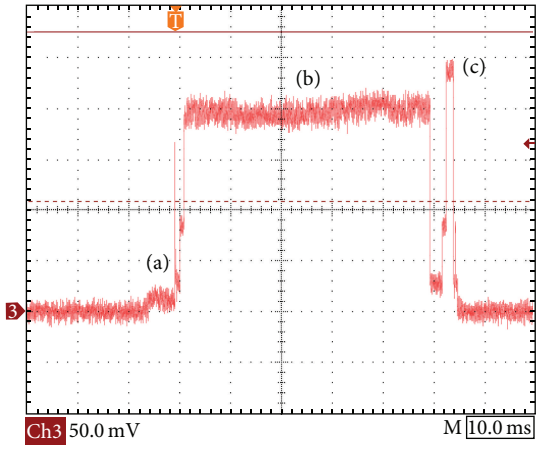


Figure 2: Consumption of a typical cycle based on a study of ODMAC [32]. The current drain is obtained by dividing the shown voltage by the shunt resistor's value (10 Ω). The activity cycle consists of the following actions: (a) sensing and packet generation, (b) waiting for a beacon from the receiver, and (c) transmitting the packet. Power consumption is dominated by the time the radio spends waiting for a beacon, that is, idle listening.

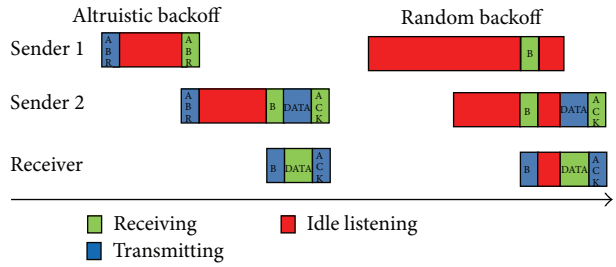


Figure 3: Altruistic versus random backoff. In RB the inevitable collision is resolved after the beacon transmission, while both nodes waste energy in idle listening waiting for it. AB uses control packets (ABR) to resolve the inevitable collision before the beacon allowing the nodes to back off earlier and save energy by decreasing the time they spend in idle listening.

forward traffic to neighbors that happen to be awake at the same time.

The widely adopted RB has its roots in avoiding long concurrent transmissions that decrease the network throughput. WSN are typically low-traffic networks with small frames that give priority to energy-efficiency rather than throughput. Our motivation for AB lies on our previous work on receiver-initiated MAC protocols, which indicates that during a transmission cycle most of the energy is consumed in idle listening, waiting for a beacon to establish a connection [32]. The cost of the data transmission itself is insignificant compared to idle listening (see Figure 2). RB implies that senders that contend for the same beacon will spend a vast amount of energy waiting for the beacon and the collision will be detected and resolved only after the beacon transmission. On the other hand, AB aims to detect the inevitable collision before the beacon transmission and allow the contending nodes to back off earlier and save energy.

3. Altruistic Backoff (AB)

Altruistic backoff is a collision avoidance mechanism that detects potential collisions and avoids them before the actual beacon transmission. Upon a wake-up event, it transmits a control packet, named ABR (*altruistic backoff request*), that identifies the beacon the node is waiting for. A node that is already waiting for the same beacon and receives this packet, altruistically backs off, offering the beacon to the node that wakes up last. At the low overhead of one extra control packet transmission per data packet transmission, collisions are mitigated and idle listening is significantly reduced. Figure 3 shows an example of AB collision avoidance compared to RB.

The presented backoff scheme does not suffer from fairness issues for two reasons. First, WSNs consist of cooperative nodes that do not have incentives to overutilize the channel. Furthermore, random channel access provides similar probabilities for all nodes to use the beacon. Essentially, the beacon and thus the channel are taken by the sender that wakes up last. Therefore, random channel access guarantees long-term fairness. In other words, as long as different senders have equal opportunities to wake up last, they have equal opportunities to take the beacon. Similarly to RB, long-term fairness can be compromised if nodes do not follow the protocol. In particular, if a sender continuously retransmits an ABR, it will always get the beacon. Generally, we do not consider this a problem, because WSNs are networks of cooperative nodes that do not have incentives to favor their performance against the performance of other nodes. However, this property is a security vulnerability that can lead to denial-of-service (DoS) attacks. Off-the-shelf security protocols, such as the receiver authentication protocol (RAP) [33], can be used to authenticate control packets in an energy-efficient manner and secure the protocol against such attacks.

Beyond being a security vulnerability, this property is used for quality of service (QoS) services through traffic differentiation. Traffic differentiation is valuable in case of applications that generate traffic of different urgency (e.g., alerts versus monitoring traffic). We define two types of data packets that correspond to two traffic classes, the high-priority class and best-effort class. The priority number that defines the priority class is included in the ABR. Upon the reception of an ABR, a node compares the priority number indicated in the ABR to the priority number of the local packet it has to transmit. If and only if the local packet belongs to the high-priority traffic class and the remote packet belongs to the best-effort traffic class, the node immediately transmits a new ABR to retake the beacon, as shown in Figure 4. As a result, the priority number guarantees that ABR retransmissions occur only when a node has a higher priority than the node that currently has the beacon.

Upon a backoff event, the time of a next transmission attempt can follow different policies with respect to the importance of the data. We can consider two extremes. On one hand, the sender might attempt to transmit immediately, as recommended for traffic of high priority. On the other

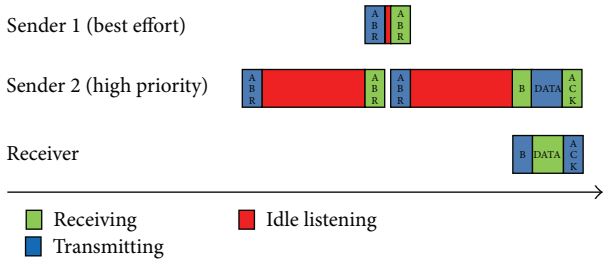


Figure 4: Traffic differentiation with AB. Nodes with traffic of high priority, upon being silenced by nodes with lower priority, immediately retransmits an ABR to retake the beacon.

hand, the sender might choose to buffer the packet and transmit it together with the following packet. We recommend this policy for best-effort traffic, as it is the policy that minimizes the energy consumption. Additionally, the sender might choose a solution between that compromises the advantages and the disadvantages of the two extremes. For the remainder of the paper and unless stated otherwise, we assume the use of the second policy.

AB is also able to support congestion control services with no additional overhead. Such feature has particular interest in adaptive protocols, such as ODMAC [18], that regulate the generated traffic to the energy resources of individual nodes. For example, consider a solar energy harvesting scenario where the receiver is placed in shadow (thus, unable to receive and forward much traffic) and the senders are placed in direct sunlight (thus, able to generate much traffic). In this scenario, the receiver would generate beacons at a low frequency and the senders would exchange many ABR frames, while contending for these few beacons. Senders may interpret frequent ABR frames, as a signal that the channel is congested and reduce the rate they generate data to avoid flooding the receiver.

4. Evaluation of AB through Simulations

In this section, we evaluate the proposed collision avoidance mechanism, AB, by comparing it with RB. The key difference between the two mechanisms lies in the way the collision is avoided. Having energy efficiency as our metric of interest, we focus the comparison on how much time the nodes spend on idle listening. In the case of AB, idle listening is the time a sender waits for a beacon. In the case of RB, idle listening is the time a sender waits for a beacon plus the number of timeslots it waits afterwards. We consider two variations of RB, namely, *constant backoff* (CB) and *binary exponential backoff* (BEB). In CB, the CW is fixed to a constant value (cw). In BEB, CW follows the binary exponential approach and cw represents the minimum contention window (CW_{min}).

We model and simulate the two methods as follows. We consider one single receiver that transmits beacons at a set frequency and a set of n nodes that are using these beacons to send their data. A round consists of the time between two beacon transmissions. Every round, each node has a probability to generate data that is equal to the ratio of the

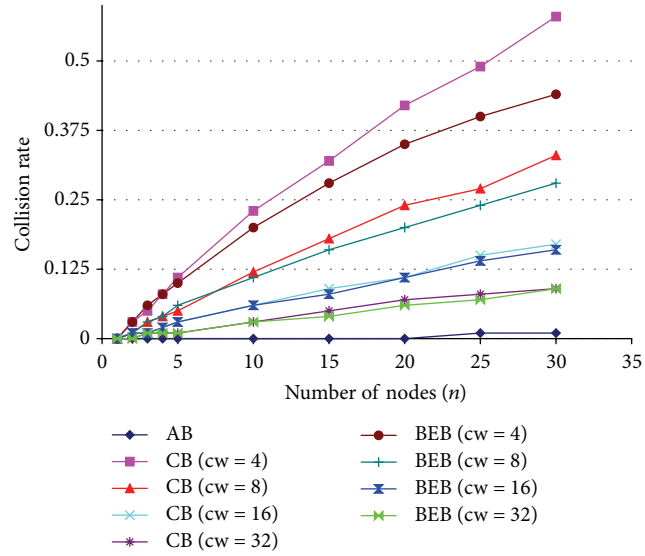


Figure 5: Collision rate for AB and random backoff with constant or binary exponential CW. In the case of BEB, cw represents the minimum contention window.

beaconing period of the receiver over its local sensing period. Nodes with data wake up at a random time during the round. The time from the beginning of a wake-up event until the reception of the following beacon or ABR is considered idle listening. In the case of AB, a collision happens when two nodes transmit the ABR at the same time frame. In the case of RB, a collision happens when two or more senders select the same and lowest random number. We set the duration of the timeslot at 100 μs and the maximum CW_{max} at 64. The simulations are conducted in MATLAB.

At the beginning we fix the beaconing period of the receiver (BP) to 4 seconds and the transmission attempt period of the receivers (SP) to 20 seconds. As a result, an average of 1/5 of the nodes in the network are contending for the channel in each round. Figure 5 shows the collision rate of the different schemes (calculated after 10000 rounds). BEB is preventing more collisions than CB for low contention windows (cw), but the difference decreases as the cw increases. This happens because as the cw increases, the probability of two or more nodes selecting the same random number decreases and, as a result, the need to double the contention window decreases. The same phenomenon appears when the number of nodes is low. AB appears more able to avoid collisions. This happens because of the random channel access. In other words, a collision can happen only if two or more nodes send an ABR simultaneously. Therefore, the time between two sequential beacons acts equivalently to a very large contention window. As a result, we expect that as we increase the contention window, the performance of BEB and CB will approach the performance of AB.

Figure 6 shows the average idle listening per transmission attempt on the same simulation. Notice that CB and BEB show a constant behavior that does not increase with neither the number of nodes nor with the contention window. The average idle listening is equal to half the period of beaconing

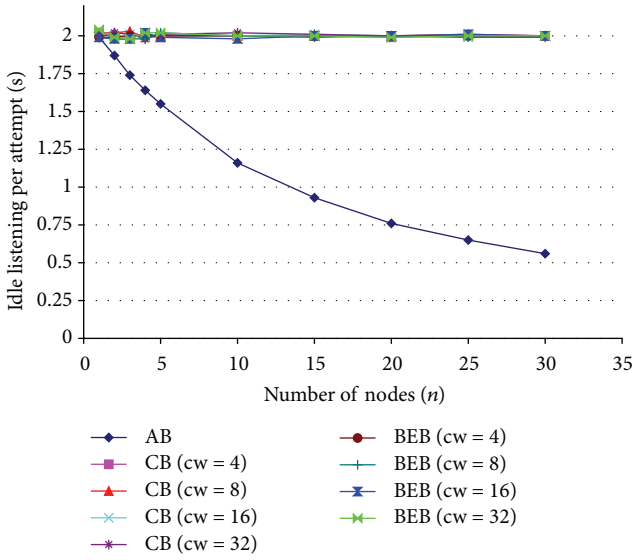


Figure 6: Average Idle listening per transmission attempt for AB and random backoff with constant or binary exponential CW. In the case of BEB, cw represents the minimum contention window.

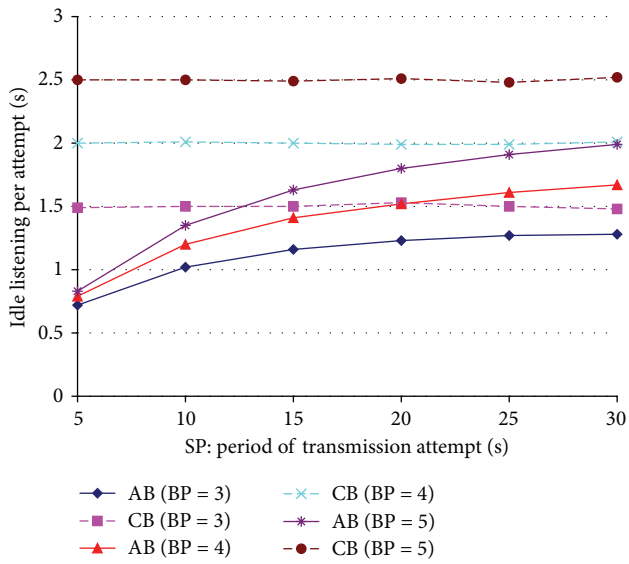


Figure 7: Average Idle listening per transmission attempt for AB and random backoff with constant CW. BP represents the beaconing period of the receiver in seconds. SP represents the period of a transmission attempt.

(BP/2). Intuitively, we expect the idle listening to increase as the contention window increases, as the number of timeslots is expected to be higher. However, the results indicate that the impact of increasing the contention window is insignificant. This behavior is explained by the size of the timeslot (100 μs) which is several orders of magnitude lower than the expected time a sender waits for a beacon. The figure shows that, in the case of AB, the average time the sender spends in idle listening decreases as the number of nodes increases. The more contention, the more ABR frames are transmitted and the faster contending nodes back off. Notice that the average idle listening for AB becomes half the period of beaconing (BP/2) when there is no contention ($n = 1$).

The above results indicate that to study idle listening is sufficient to consider only one version random backoff. In Figure 7, we consider 5 contending senders and CB with fixed contention window (cw = 4). We variate the period of a transmission attempt (SP) and the period of beaconing (BP). The results show a similar constant behavior for CB, while the average idle listening of AB decreases as the traffic increases (SP decreases).

Figure 8 shows the distribution of successful transmissions over all the contending nodes, considering $n = 20$, BP = 4 s, and SP = 20 s, for the case of AB. We can observe that random channel access leads to equal probabilities for every node to be the last sender to wake up before the beacon. Therefore, AB provides long-term fairness for channel access.

Next, we demonstrate the ability of AB to differentiate traffic to provide QoS. We consider two classes of traffic, namely, *high priority* and *best effort*. Sensor nodes mark the data that they generate as *high priority* with a probability P. According to the protocol specification, the sensor node that wakes up last and has a data packet marked as *high priority* takes the beacon. If there is no sensor node with *high priority*

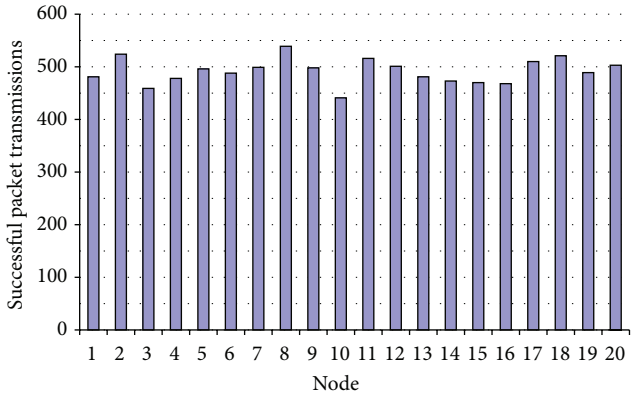


Figure 8: The distribution of successful transmissions indicates that AB provides long-term fairness, as contending nodes have equal probabilities to use the channel.

data packets contending for the medium, the sensor node that wakes up last and has a data packet marked as *best effort* takes the beacon. For the following simulation, we consider $P = 0.05$, BP = 1 s, and SP = 3 s. Figure 9 shows the average ratio of the amount of data packets that take a beacon over the total amount of generated packets, for each priority class (calculated after 10000 rounds). As the contention increases, a larger amount of *best effort* traffic backs off, while the *high priority* traffic is less affected. Essentially, AB sacrifices less important traffic to prioritize urgent traffic. The slight decreasing trend for the *high priority* traffic is attributed to the rounds that multiple nodes with *high priority* traffic contend with each other.

The results indicate that AB is long-term fair and scales well with high contention, as the ABR frames efficiently put

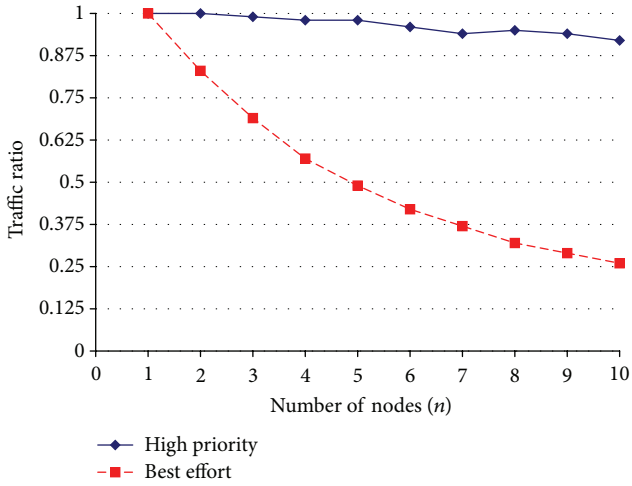


FIGURE 9: The average ratio of the amount of data packets that take a beacon over the total amount of generated packets for each priority class. As the contention increases, the protocol sacrifices *best effort* traffic for *high priority* traffic.

the contending nodes to sleep early and less energy is wasted in idle listening. Furthermore, AB is able to differentiate traffic to provide QoS.

5. Implementation of AB

We implemented AB as an extension to the implementation of on demand MAC (ODMAC) [32] for Texas Instruments' eZ430-rf2500 nodes [14]. The nodes consist of an MSP430 microcontroller (MCU) and a CC2500 radio, operating in the 2.4 GHz band. ODMAC is a receiver-initiated MAC protocol for energy harvesting-wireless sensor networks (EH-WSNs) that has been developed upon the principle of dynamically adapting the duty cycles to the amount of harvested energy. Nevertheless, the basic communication scheme of ODMAC follows the receiver-initiated paradigm of communication, as described in Section 2. We refer the reader to [32] for the details of implementation of the protocol. ODMAC is implemented as a finite state machine (FSM). Its functionality is mainly based upon two routines, namely, *send* and *receive*. Unless one of these two handlers is invoked, ODMAC is in sleeping state and the radio is off. The *send* routine generates and formats a packet around the payload (i.e., the result of a sensing operation). When the packet is ready, the radio is switched on into listening mode and the state machine awaits for an interrupt signaling the reception of an appropriate beacon. Should this happen, ODMAC continues its execution and the data packet is transmitted. At the end of a packet transmission, the radio is switched off.

AB extends the ODMAC *send* routine as follows. After the packet generation, an ABR frame that includes information about the intended receiver is generated. After a successful CCA the transmission of ABR follows. Then, the radio is switched to listening mode and the sender begins to listen for a beacon. Listening is interrupted either by the reception of the expected beacon or by the reception of an ABR that indicates interest for the same beacon. In the former case, data transmission follows normally. In the latter case, the routine returns and indicates a backoff. It should be noted that the *send* routine performs one attempt to transmit the packet. In case of backoff, the higher layer is free to decide at which point in the future will attempt again to transmit the same packet. The state machine in Figure 10 summarizes the operation of AB as part of the ODMAC protocol.

For the traffic differentiation services of AB, we extend the implementation by adding a priority bit in the header of ABR control packets. The priority bit indicates if the data packet is classified as *high priority* or *best effort*. When a sender that waits for a beacon receives another ABR packet, it compares its local priority bit with the received priority bit. If and only if the local data packet is classified as *high priority* and the received ABR indicates a *best effort* data packet, the sender retakes the channel by invoking the *send* routine again.

6. Experimental Results

In this section, we experimentally evaluate AB. For the purposes of a comparison with RB, we also implemented a simple variation of the protocol with constant contention window, CB, in the ODMAC protocol. We chose to implement the CB variation because our simulations (see Section 4) indicate that the variation of the protocol and the length of the contention window do not affect the idle listening overhead significantly. CB is implemented by adding a random delay between the reception of a beacon and the transmission of the data. In particular, we use a constant contention window (cw = 4) and a timeslot of 100 MCU cycles (≈100 μs).

To measure the idle listening time interval, we use the internal timer unit, which is set to use the low frequency oscillator (12 KHz) that remains active when the MCU goes into low power (i.e., sleeping) modes. Because of the size of its counter register (16 bits), the timer is able to measure time intervals up to approximately 5.5 seconds. Each node is set to keep the sum of all the time it spent in idle listening since reset and reports the value in every data packet. In addition to that, a sequence number of all the data transmission attempts are also reported. Using the two aforementioned values, we can estimate the average time a node spent in idle listening per data transmission attempt.

For the experiments presented in this section, we use the following test bed. We use a single-hop star topology with a set of senders contending to transmit to a single receiver. The contending senders are placed physically close to each other and to the receiver, in order to mitigate any packet losses due to channel errors. The receiver is connected to a laptop, through which we collect all the received packets. The receiver node is set to transmit beacons but never generate data of its own. A set of sender nodes are configured to periodically transmit data to the receiver. We use ODMAC's randomization feature [32] to randomize the period of data transmission attempts and enforce random channel access. In particular, after each data transmission, the wake-up interrupts are randomized over the whole space of the register. The node then calls the *send* routine once every *sm* wake-up

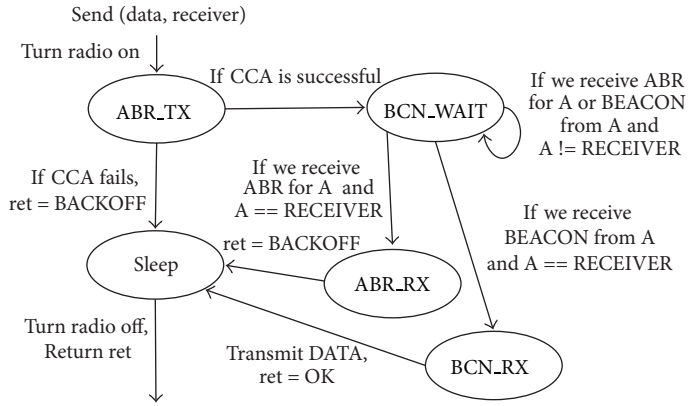


FIGURE 10: The finite state machine that specifies the operation of AB as part of the ODMAC protocol.

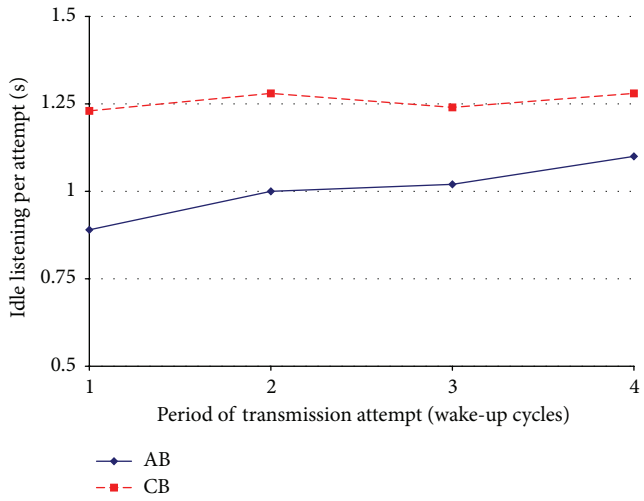


FIGURE 11: Average idle listening per transmission attempt for ab and random backoff with constant CW. Wake-up interrupts are uniformly randomized after each transmission to enforce random channel access.

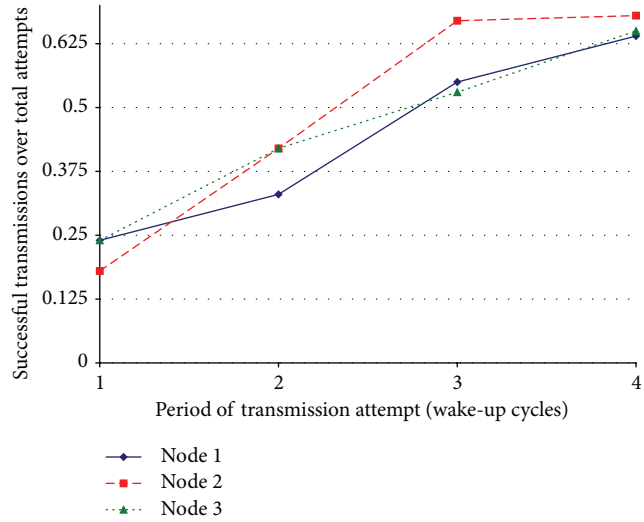


FIGURE 12: The ratio of successful transmissions over the total number of transmission attempts indicates that AB is long-term fair.

interrupts, where *sm* is a configurable parameter that controls the average period of data transmission attempts.

In the experiment shown in Figure 11, we set the beaconing period of the receiver to 4 seconds and we used 3 contending senders. In the x-axis we variate the period of a transmission attempt for all the senders in wake-up interrupts, that is, the *sm* parameter. The duration of each experiment was 1 hour. The results indicate a similar trend to the respective simulation experiment, shown in Figure 7, which verifies the energy consumption improvements of AB. CB follows a similar constant behavior. AB, on the other hand, is spending less time in idle listening as the traffic increases. Figure 12 shows the ratio of successful transmissions over the total number of transmission attempts for the same experiments for AB. The results demonstrate the long-term fairness of the protocol, as the nodes appear to have equal opportunities to take the channel. We can notice that none of the senders is led to starvation and the number of times they took the channel is at the same order of magnitude between the three nodes. The relative difference between the senders is attributed to the duration of the experiment (1 hour). We expect longer experiments to smooth such differences out.

In the next experiment, we fix the period of transmission attempts to 2 wake-up cycles and we variate the number of contending nodes from 1, that is, no contention, to 4. Figure 13 shows the average time each node spends on idle listening per transmission attempt for the two protocols. The duration of each experiment was 1 hour. The results follow a similar trend to the respective simulation experiment, shown in Figure 6. In particular, when there is no contention, the two protocols have similar performance. For the case of CB, the average time spent in idle listening remains constant, being dominated by the time the node waits for a beacon. In the case of AB, on the other hand, idle decreases as the contention increases.

Next, we evaluate the long-term fairness of AB in the scenario of contending senders with different traffic generation

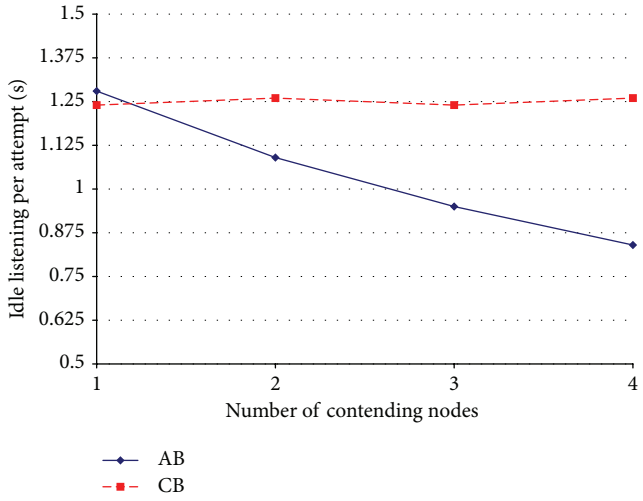


FIGURE 13: Average idle listening per transmission attempt for ab and random backoff with constant CW for different numbers of contending nodes.

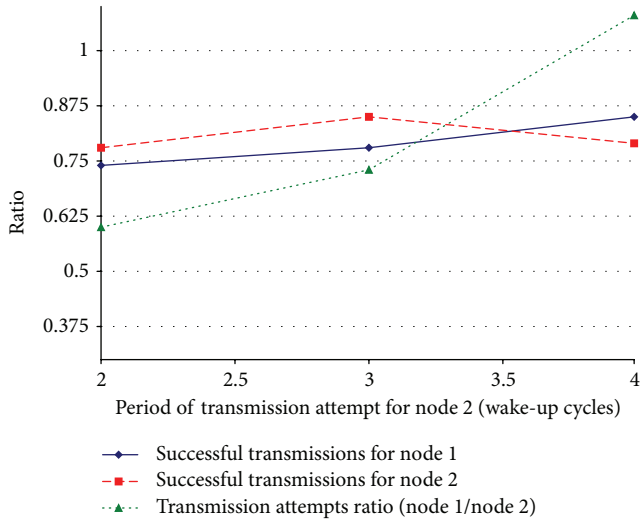


FIGURE 14: The ratio of successful transmissions over the total number of transmission attempts for node 1 and node 2. The period of transmission attempts for node 1 is fixed to 4 wake-up cycles. The triangle-line shows the ratio of the packets generated by node 1 over node 2.

frequencies. Such scenario has interest in cases of nodes with different forwarding duties or different power resources (e.g., energy harvesting sensor nodes have access to different levels of ambient energy). The experiment is designed as follows. We use 2 nodes and fix the period of transmission attempts of the first node to 4 wake-up interrupts, while varying the period of the second node from 2 to 4. The duration of each experiment is 2 hours. Figure 14 shows the results of the experiment. The triangle-line shows the ratio of the packets generated by node 1 over node 2, which increases as the period of transmission attempt of node 2 increases. Note that, when the nodes have equal periods, the ratio is close to 1. We observe that, despite the fact that the two nodes attempt to

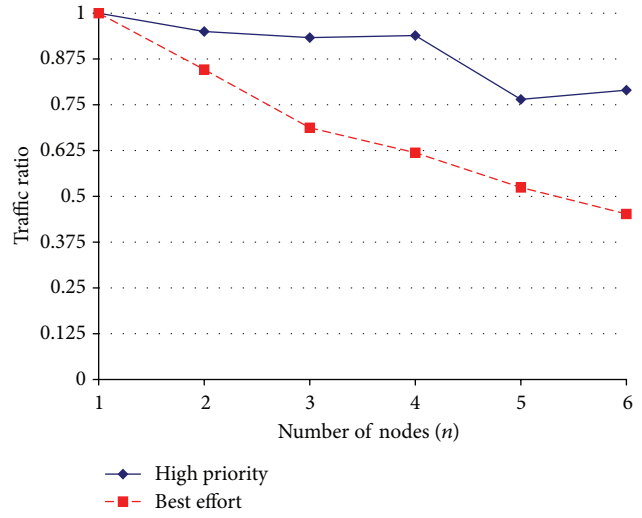


FIGURE 15: The average ratio of the amount of data packets that take a beacon over the total amount of generated packets for each priority class. As the contention increases, the protocol sacrifices *best effort* traffic for *high priority* traffic.

use the channel at different frequencies, they maintain equal opportunities to obtain the beacon. The ratio of success full packet transmissions over the total amount of transmission attempts shows a constant behavior.

In the next experiment, we experimentally evaluate traffic differentiation by replicating the simulation shown in Figure 9. The beaconing period of the receiver is set to 1 second and the period of transmission attempts of the senders is randomized with an average of approximately 3 seconds. Moreover, nodes generate *high priority* data packets with a probability of $P = 0.05$. Figure 15 shows the average ratio of the amount of data packets that take a beacon over the total amount of generated packets, for each priority class. The duration of each experiment is 1 hour. Due to hardware constraints, the experiment was conducted with up to 6 contending nodes. The results validate the simulations and show that as the contention increases, a larger amount of *best effort* traffic backs off, giving priority to the *high priority* traffic.

In the last figure, we validate the simulations by comparing their estimations to the results obtained through the experimental evaluation. In particular, we configure the simulator to the exact same configuration that is used in the test-bed experiment presented in Figure 13. In the experiment the period of transmission attempts of the senders is set to 2 wake-up cycles that are uniformly randomized over the whole space of the register, leading to an average period of approximately 5.5 seconds. Thus, in the simulator we set period of transmission attempts to 5.5 seconds. The beaconing period of the receiver is set to 3 seconds. Figure 16 plots the ratio of the average idle listening per transmission attempt of AB over CB as obtained from the simulation and the test-bed experiment. Observe that both simulations and test-bed experiments give close results, while the behavior of the protocol follows the same trend. The difference indicates that, in the experiments, random access is not as uniformly

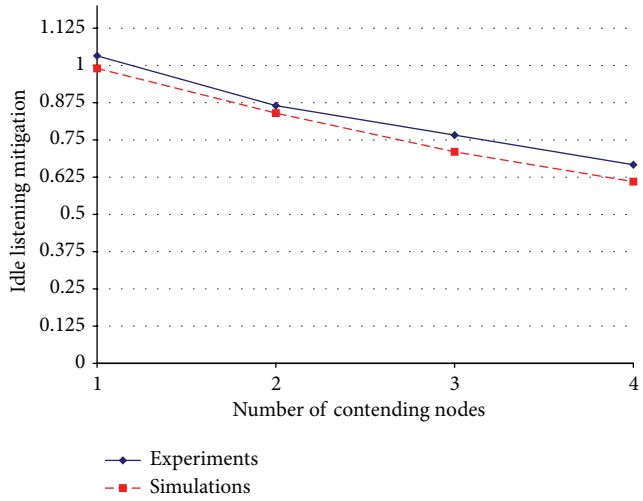


FIGURE 16: The ratio of the average idle listening per transmission attempt of AB over CB as obtained from the simulations and the test-bed experiments.

distributed throughout the interval between two beacons, as assumed in the simulations.

The results of the experiments verify the trends that are suggested by the simulations, presented in Section 4. AB scales well with both high contention and high traffic and provides equal opportunities for the contending nodes to access the channel. Detecting the inevitable collisions before the beacon transmission allows the nodes to resolve the collision before significant amount of energy is wasted in idle listening while waiting for the beacon.

7. Conclusion

In this paper, we have focused on receiver-initiated MAC protocols in wireless sensor networks. Such protocols initiate the data exchange with a beacon that is transmitted by the receiver and states its availability to receive traffic. Beacons nullify the benefits of random channel access, as they constitute points of potential collisions even in situations of sparse traffic. We have proposed AB, a collision avoidance mechanism that exploits random channel access to avoid collisions while decreasing the time nodes waste energy in idle listening. Simulations and experiments indicate that AB is long-term fair and scales well with increasing levels of contention. Furthermore, AB provides QoS by prioritizing traffic of different urgencies. AB is compared to the commonly used collision avoidance mechanism, namely, random backoff, and the results demonstrate the energy savings that can be achieved with AB. Finally, we have discussed an implementation of the proposed collision avoidance mechanism for Texas Instruments' eZ430-rf2500 sensor nodes [14], incorporated in the ODMAC protocol [32].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] A. Bachir, M. Dohler, T. Watteyne, and K. K. Leung, "MAC essentials for wireless sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 2, pp. 222–248, 2010.

[2] IEEE, "IEEE std. 802.15.4-2003: wireless medium access control (mac) and physical layer (phy) specifications for low rate wireless personal area networks (lr-wsns)," 2003.

[3] B. Otal, L. Alonso, and C. Verikoukis, "Design and analysis of an energy-saving distributed mac mechanism for wireless body sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, article 10, 2010.

[4] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications (INFOCOM '02)*, vol. 3, pp. 1567–1576, June 2002.

[5] T. van Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys '03)*, vol. 5–7, pp. 171–180, ACM, November 2003.

[6] P. Lin, C. Qiao, and X. Wang, "Medium access control with a dynamic duty cycle for sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '04)*, vol. 3, pp. 1534–1539, IEEE, March 2004.

[7] A. Antonopoulos and C. Verikoukis, "Network-coding-based cooperative ARQ medium access control protocol for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 601321, 9 pages, 2012.

[8] A. El-Hoiydi and J.-D. Decotignie, "WiseMAC: an ultra low power MAC protocol for multi-hopWireless sensor networks," in *Proceedings of the 1st International Workshop on Algorithmic Aspects of Wireless Sensor Networks (ALGOSENSORS '04)*, pp. 18–31, 2004.

[9] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proceedings of the 2nd ACM International Conference on Embedded Networked Sensor Systems (SenSys '04)*, pp. 95–107, ACM, November 2004.

[10] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks," in *Proceedings of the 4th ACM International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 307–320, ACM, November 2006.

[11] C. Cano, B. Bellalta, A. Sfairopoulou, and M. Oliver, "Low energy operation in WSNs: a survey of preamble sampling MAC protocols," *Computer Networks*, vol. 55, no. 15, pp. 3351–3363, 2011.

[12] E.-Y. A. Lin, J. M. Rabaey, and A. Wolisz, "Power-efficient Rendez-vous schemes for dense wireless sensor networks," in *Proceedings of the IEEE International Conference on Communications (ICC '04)*, vol. 7, pp. 3769–3776, June 2004.

[13] Y. Sun, O. Gurewitz, and D. B. Johnson, "RI-MAC: a receiver-initiated asynchronous duty cycle MAC protocol for dynamic traffic loads in wireless sensor networks," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 1–14, ACM, November 2008.

[14] Texas Instruments, "eZ430-RF2500 Development Tool," SLAU227E, 2009, http://www.ti.com/lit/ug/slau227e/slau227e.pdf.

[15] P. Dutta, S. Dawson-Haggerty, Y. Chen, C.-J. M. Liang, and A. Terzis, "A-MAC: a versatile and efficient receiver-initiated

link layer for low-power wireless," *ACM Transactions on Sensor Networks*, vol. 8, no. 4, article 30, 2012.

[16] Y.-T. Yong, C.-O. Chow, J. Kanesan, and H. Ishii, "EE-RI-MAC: an energy-efficient receiver-initiated asynchronous duty cycle MAC protocol for dynamic traffic loads in wireless sensor networks," *International Journal of Physical Sciences*, vol. 6, no. 11, pp. 2633–2643, 2011.

[17] L. Tang, Y. Sun, O. Gurewitz, and D. B. Johnson, "PW-MAC: an energy-efficient predictive-wakeup MAC protocol for wireless sensor networks," in *Proceedings of the 30th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '11)*, pp. 1305–1313, IEEE, April 2011.

[18] X. Fafoutis and N. Dragoni, "ODMAC: an on-demand mac protocol for energy harvesting—wireless sensor networks," in *Proceedings of the 8th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN '11)*, pp. 49–56, ACM, November 2011.

[19] Y. Peng, Z. Li, D. Qiao, and W. Zhang, "Delay-bounded MAC with minimal idle listening for sensor networks," in *Proceedings of the 30th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '11)*, pp. 1314–1322, April 2011.

[20] Y. Sun, O. Gurewitz, S. Du, L. Tang, and D. B. Johnson, "ADB: an efficient multihop broadcast protocol based on asynchronous duty-cycling in wireless sensor networks," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09)*, pp. 43–56, ACM, November 2009.

[21] P. Yadav and J. A. McCann, "YA-MCA: handling unified unicast and broadcast traffic in multi-hop wireless sensor networks," in *Proceedings of the 7th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS '11)*, pp. 1–9, IEEE, June 2011.

[22] J. Li, D. Zhang, and L. Guo, "DCM: a duty cycle based multi-channel MAC protocol for wireless sensor networks," in *Proceedings of the IET International Conference on Wireless Sensor Network (IET-WSN '10)*, pp. 233–238, November 2010.

[23] L. Tang, Y. Sun, O. Gurewitz, and D. B. Johnson, "EM-MAC: a dynamic multichannel energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the 12th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '11)*, p. 23, ACM, May 2011.

[24] X. Fafoutis and N. Dragoni, "Analytical comparison of MAC schemes for energy harvesting—wireless sensor networks," in *Proceedings of the 9th International Conference on Networked Sensing Systems (INSS '12)*, IEEE, June 2012.

[25] J. Rousselot, A. El-Hoiydi, and J.-D. Decotignie, "WideMac: a low power and routing friendly MAC protocol for ultra wideband sensor networks," in *Proceedings of the IEEE International Conference on Ultra-Wideband (ICUWB '08)*, vol. 3, pp. 105–108, September 2008.

[26] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Standards Association Std., 2012.

[27] K. Whitehouse, A. Woo, F. Jiang, J. Polastre, and D. Culler, "Exploiting the capture effect for collision detection and recovery," in *Proceedings of the 2nd IEEE Workshop on Embedded Networked Sensors (EmNetS-II '05)*, pp. 45–52, IEEE, May 2005.

[28] P. Huang, C. Wang, L. Xiao, and H. Chen, "RC-MAC: a receiver-centric medium access control protocol for wireless sensor networks," in *Proceedings of the IEEE 18th International Workshop on Quality of Service (IWQoS '10)*, pp. 1–9, June 2010.

[29] D. Kominami, M. Sugano, M. Murata, and T. Hatauchi, "Energy-efficient receiver-driven wireless mesh sensor networks," *Sensors*, vol. 11, no. 1, pp. 111–137, 2011.

[30] Q. Lampin, D. Barthel, I. Augé-Blum, and F. Valois, "SARI-MAC: the self adapting receiver initiated MAC protocol for wireless sensor networks," in *Proceedings of the IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '12)*, pp. 12–18, IEEE, October 2012.

[31] X. Wang, X. Zhang, G. Chen, and Q. Zhang, "Opportunistic cooperation in low duty cycle wireless sensor networks," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, pp. 1–5, IEEE, May 2010.

[32] X. Fafoutis, A. D. Mauro, and N. Dragoni, "Sustainable medium access control: implementation and evaluation of ODMAC," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '13)*, IEEE, 2013.

[33] A. D. Mauro, X. Fafoutis, S. M. Mödersheim, and N. Dragoni, "Detecting and preventing beacon replay attacks in receiver-initiated MAC protocols for energy efficient WSNs," in *Proceedings of the 18th Nordic Conference on Secure IT Systems (NordSec '13)*, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 480372, 17 pages
http://dx.doi.org/10.1155/2014/480372

Research Article

A QoS-Based Wireless Multimedia Sensor Cluster Protocol

Juan R. Diaz,[1] Jaime Lloret,[1] Jose M. Jimenez,[1] and Joel J. P. C. Rodrigues[2]

[1] *Universidad Politécnica de Valencia, Camino Vera s/n, 46022 Valencia, Spain*
[2] *Instituto de Telecomunicações, University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal*

Correspondence should be addressed to Jaime Lloret; jlloret@dcom.upv.es

Received 15 January 2014; Accepted 15 March 2014; Published 18 May 2014

Academic Editor: Sana Ullah

Wireless Sensor Networks (WSNs) provide a wireless network infrastructure for sensed data transport in environments where wired or satellite technologies cannot be used. Because the embedded hardware of the sensor nodes has been improved very much in the last years and the number of real deployments is increasing considerably, they have become a reliable option for the transmission of any type of sensed data, from few sensed measures to multimedia data. This paper proposes a new protocol that uses an ad hoc cluster based architecture which is able to adapt the logical sensor network topology to the delivered multimedia stream features, guaranteeing the quality of the communications. The proposed protocol uses the quality of service (QoS) parameters, such as bandwidth, delay, jitter, and packet loss, of each type of multimedia stream as a basis for the sensor clusters creation and organization inside the WSN, providing end-to-end QoS for each multimedia stream. We present real experiments that show the performance of the protocol for several video and audio cases when it is running.

1. Introduction

The number of Wireless Sensor Network (WSN) real deployments is increasing considerably in the last years [1, 2], mainly because of their huge benefits [3]. New wireless technology standards, recent advances in energy-efficient hardware and video coding algorithms are allowing multimedia delivery over ad hoc networks.

Nowadays, the features of the sensor nodes and smart devices are very similar to the regular personal computer features. Last generation of sensor nodes can include advanced models of CPUs, with several cores, 1 or 2 GB of RAM, and storage capacities up to 64 GB. Moreover, they can include multiple wireless interfaces such as Bluetooth, Wi-Fi, 3G, and 4G,

The amounts of multimedia services that can be offered through the network are very large [4, 5]: VoIP, IPTV, radio, teaching, multimedia streaming, games, and so forth. There are a lot of multimedia platforms and protocols used in different fields [6], from the entertainment to the training in the business environment.

Ad hoc network is a self-organizing multihop system of wireless nodes which can communicate with each other without a preexisting infrastructure [7]. Multimedia ad hoc networks can be ideal to allow a distributed multimedia service in commercial and social environments that require high visibility to the offered products.

With the widespread use of wireless technology, the ability of mobile wireless ad hoc networks to support multimedia services with quality of service (QoS) has become a challenging research subject as described by Khoukhi and Cherkaoui in [8]. According to Barenco Abbas et al. the main goal of QoS is to achieve a more deterministic network behavior [9]. Chen et al. suggested in [10] that when we need to provide an acceptable QoS in the network, we should define the values of QoS metrics in order to establish the necessary requirements. These requirements are different if it is a real-time service or an on-demand service.

Due to the severe limitations of the ad hoc networks (in terms of energy, processing power, memory, bandwidth, etc.), it is necessary to carefully design the multimedia ad hoc network protocol. Some works are focused on proposing multichannel cross-layer architectures [11], while others are focused on providing fast rerouting algorithms [12], but in this case we focus our research on providing the best topological structure based on the type of multimedia streams.

There have been many studies proposing different topological structures for ad hoc networks that can be summarized into two main types: planar and hierarchical topologies [13]. Planar topologies in ad hoc networks may be of great complexity, mainly in mobile ad hoc networks, because any node displacement may change the entire network topology. For this reason most of researchers have proposed the use of a hierarchical structure for performing an ad hoc network topology [14]. In many cases this hierarchical structure split nodes into different groups called clusters [15, 16].

In this paper we show the design and performance test of a new multimedia protocol which takes into account the QoS in WSNs. The protocol uses the QoS parameters to structure the network topology. Then, a node decides where to join based on its QoS needs. The protocol is based on the architecture proposed by us for wireless ad hoc networks in [17]. While in [17] we only propose architecture for ad hoc networks, in this paper we have particularized the architecture to Wireless Sensor Networks and we have focused this work to the design and deployment of the network protocol.

This paper is organized as follows. Section 2 presents the research papers related with our work. Proposed protocol and architecture are described in detail in Section 3. The system operation is explained in Section 4. Section 5 shows the obtained results and our discussion of the performance study. Finally, in Section 6, conclusion and future work are shown.

2. Related Work

We have structured the related work section in 2 parts. The first part shows several cluster formation algorithms, while the second part discusses published cluster-based multimedia ad hoc networks.

There are several surveys that review existing works on cluster formation algorithms. On one hand, according to Wei and Anthony Chan [18] cluster topologies can be classified into four categories: single-hop or multihop, stationary or mobile, synchronous or asynchronous, and location-based or non-location-based. On the other hand, Yu and Chong [19] made a categorization of clustering schemes in stationary and mobile ad hoc networks and sensor. They classified 14 proposed clustering schemes into six categories based on their main objectives. Moreover, they discussed each clustering scheme in terms of objective, mechanism, performance, and application scenario and discussed the similarities and differences between schemes of the same clustering category. We have also found [20], authored by Agarwal and Motwani. They reviewed several clustering algorithms which help organize mobile ad hoc networks in a hierarchical manner and presented their main characteristics. With this survey we see that a cluster-based MANET has many important issues to examine, such as the cluster structure stability, the control overhead of cluster construction and maintenance, the energy consumption of mobile nodes with different cluster-related status, the traffic load distribution in clusters, and the fairness of serving as cluster heads for a mobile node.

We have also found two papers written by Abbasi and Younis [21] and Boyinbode et al. [22], which present a synthesis of existing clustering algorithms in WSNs and highlight the challenges in clustering. They survey different clustering algorithms for WSNs, emphasizing their objectives, features, complexity, and so forth. They also compare their metrics such as convergence rate, cluster stability, cluster overlapping, location awareness, and support for node mobility.

Despite this review, we would like to mention 4 clustering algorithms not included in these surveys because of their importance.

Ramachandran et al. [23] proposed two new distributed clustering algorithms for wireless ad hoc networks. They presented a 2-stage O(N) randomized algorithm for a N node complete network, which finds the minimum number of star-shaped clusters, all at their maximum size. They also proved the correctness of this algorithm. They then presented a completely deterministic O(N) algorithm in which cluster heads are elected autonomously by the nodes. They compared their performance using simulations on top of Bluetooth's device discovery procedures. Results show that the randomized algorithm performs better with respect to both cluster and network formation times.

Chatterjee et al. [24] proposed a weight based distributed clustering algorithm (WCA) which can dynamically adapt itself with the ever changing topology of ad hoc networks. Their approach restricts the number of nodes to be catered by a cluster head so that it does not degrade the MAC functioning.

Lehsaini et al. [15] showed the development of an architecture that creates clusters and establishes connections between sensors of the same type by building different sensor networks. In their proposal the cluster heads manage the network since they have connections with other cluster heads and these connections allow connecting cluster members from different clusters when they are of the same type, forming a specialized network. One of the main goals is that if all cluster heads switch off at the same time, the system is able to continue working, although there will not be new connections between clusters through the cluster heads.

Kavitha and Karthikeyan [25] proposed an energy enhanced version of the M-SPIN (EEM-SPIN) protocol using WCA for WSNs. It has the flexibility of assigning different weights and takes into account combined metrics to form clusters automatically. Limiting the number of nodes inside a cluster allows restricting the number of nodes catered by a cluster head so it does not degrade the MAC functioning. For a fixed cluster head election scheme, a cluster head with constrained energy may drain its battery quickly due to heavy utilization. In order to spread the energy usage over the network and achieve a better load balancing among cluster heads, reelection of the cluster heads may be a useful strategy.

Next, we review how some of the main cluster-based multimedia ad hoc networks are created.

Huang et al. [26] have presented a cluster-based model to support multimedia service. The proposed model transmits multimedia streaming stably in ad hoc networks, while mobile users who consume multimedia streams tend towards

group-based behavior. An on-demand connection prediction to measure the likelihood of connectivity of cluster-based routes in a future time is applied to the cluster-based transmission of multimedia streaming. They proposed a routing method called PLCBRP (cluster-based routing with the prediction of connection probability), which combines the cluster-based routing protocol with the prediction scheme. PLCBRP discovers an optimal loosely cluster-based route for transmitting long multimedia streams. Simulation results indicate that PLCBRP delivers more data packets and provides more quality on the transmission of multimedia streaming than other flat on-demand routing protocols do.

Tang and Li [27] developed a QoS supporting scheme for dynamic traffic conditions by controlling data generating rates at individual clusters. Besides, they have investigated an explicit solution on the energy distribution at different clusters in the WSN, based on an optimal energy allocation criterion. The obtained network energy distribution formula is particularly convenient for node deployment design in WSNs. The proposed algorithm is presented and validated by numerical simulations. Some situations are also discussed and presented by experimental examples.

Rosário et al. proposed MEVI in [28] a smart multihop hierarchical routing protocol for efficient video communication over Wireless Multimedia Sensor Networks. It combines a cluster formation scheme with low signaling overhead in order to ensure reliable multihop communication between cluster heads and base stations. For route selection, a cross-layer solution selects routes based on network conditions and energy issues and a smart scheme to trigger multimedia transmission according to sensed physical environmental conditions. The cluster approach aims to minimize the energy consumption. MEVI allows the transmission of multimedia content with QoS/QoE support by introducing a hierarchical routing protocol. Simulation experiments show the benefits of MEVI in disseminating video content for large and small field size, compared with low-energy adaptive clustering hierarchy (LEACH) and power efficient multimedia routing (PEMuR) in terms of network lifetime and video quality level.

In [29], Diaz et al. propose a new multimedia-oriented application layer protocol, which takes into account the multimedia services offered by the nodes in the wireless ad hoc network in order to select the best multimedia service provider node and to provide the best QoE and QoS to the nodes participating in the ad hoc network. Authors show the designed protocol and decision algorithms in order to provide the best multimedia service to the end users. Video streaming is more challenging problem than audio streaming. It requires a considerable bandwidth to provide enough QoS. The system takes into account the delay, jitter, lost packets, and bandwidth parameters in order to select the best service provider node. Moreover, the system takes into account the estimated QoE parameter (based on a previously studied formula) and the closest node which implies less RTT and thus lower zapping times, in order to have the best QoE. The authors validate their proposal through an implemented study case.

The protocol proposed in this paper is based on the architecture proposed by us for wireless ad hoc networks in [17]. While our previous work was based on the architecture definition and deployment, this paper is focused on the protocol including the restriction given by the sensor networks. Moreover, we have included the case where a request can be performed from outside the WSN (like in a regular WSN) and the tests to perform the real experiments are completely different.

3. Protocol Description

In this section we are going to describe the proposed protocol. First, we describe the architecture features, the elements of the framework, and their relationship. Then, we explain the characteristics of the protocol, the structure of the protocol header, and the protocol fields. Finally, we show the messages designed for the proper operation of our proposal.

The main objective of the protocol is to let the sensor nodes communicate taking into account multimedia flow characteristics. It uses a cluster-based ad hoc architecture that will control the QoS parameters for each multimedia communication, by establishing the appropriated values and guaranteeing the service along the time. The protocol allows the sensor to communicate and exchange information about their state and properties. Moreover, sensor nodes use this information to determine the most appropriate neighbors. The protocol dynamically manages the creation of the cluster as a function of the network features, the number of devices, the sensor capacity, and the multimedia flows.

3.1. System Architecture. The starting point of our system is a set of wireless sensor nodes located in a delimited place which form a WSN. Each wireless sensor node has different power, processing, memory, and transmission capacities. They are able to select other wireless sensor nodes as ad hoc neighbors if they are under their radio coverage area. Wireless sensor nodes are responsible for retransmitting the multimedia flows, which may be audio or video, and can use a wide range of codecs.

Figure 1 shows the elements of a cluster. Some sensor nodes are able to provide sensed data to the WSN as audio IP or video IP services. There are three types of communications as a function of the source or destination of the communication: (1) communication started from outside the WSN to a node placed inside the WSN, (2) communication started from a node placed inside the WSN to a destination outside the WSN, and (3) communication started from a node placed inside the WSN to a node of the WSN. A node from an external network can provide multimedia contents and audio and video real-time communication services.

Wireless sensor nodes can sense multimedia data or act as a data forwarding nodes inside the WSN. They can communicate with other nodes under their coverage area. New nodes will select the better reachable cluster based on their features and the type of multimedia traffic that is going to be transmitted. We can distinguish in Figure 1 two types of nodes, sensor nodes that do not have any connection with nodes from an external network (they can only establish connections with nodes from their cluster) and sensor nodes

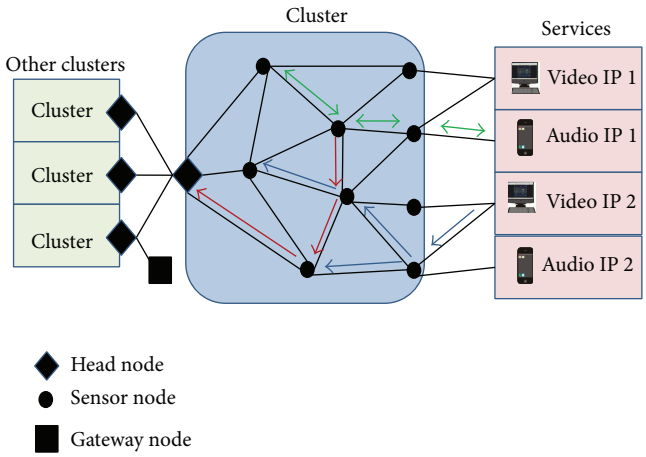


Figure 1: Cluster elements and possible communications.

that have connections with other clusters (cluster heads) or with an external network (gateway nodes). Gateway nodes have two interfaces at least in order to connect with the WSN and with the external network.

The network is organized in clusters. Every cluster of the architecture is dedicated to a specific multimedia flow, which will be identified by predefined multimedia profiles. We have created a Multimedia Init Profile (MIP) in order to manage the configuration of the sensor nodes [26]. MIP defines which type of multimedia flow can be delivered by the sensor node. MIP groups in a single logical component all required information to guarantee the adequate QoS to the multimedia traffic. MIP gathers the restrictions that will be applied to QoS parameters for the multimedia flows (Bandwidth, Delay, Jitter, and Lost Packets) and the cluster properties (maximum number of hops and the number of connections with external networks). There is only one MIP associated with each cluster in the WSN, but there could be several clusters using the same MIP. When a MIP is assigned to a sensor node, the following information is assigned: type of multimedia traffic (audio or video), range of codecs that can be used by the multimedia flows inside the cluster, maximum bandwidth available for retransmissions, and the maximum admissible Delay, Jitter, and Lost Packets.

Figure 2 shows the elements of the proposed architecture and their relationship. The architecture defines three operation levels: *Hardware Infrastructure*, *Logic Management,* and *Admin Interface*.

Hardware Infrastructure level is formed by the elements in charge of building the physical and logical network topology. The physical topology is made of wireless sensor nodes. Each node can be head node, gateway node, or sensor node (each node can only have one role). When a sensor node starts for its first time, it searches other sensor nodes in its coverage area. This process lets the node exchange the required information to group the nodes in clusters by means of the developed protocol. Then, the logical topology is created. Sensor nodes can only belong to a single cluster and have neighborhoods with the nodes of that cluster. The head node belongs to a single cluster but can have neighborhoods with other clusters' head nodes. The wireless connections between head clusters create a higher hierarchical level that allows exchanging information between clusters. The criteria used to determine which cluster will be the sensor node joined to are based on the MIP associated with the sensor node, and thus on the type of multimedia traffic it is disposed to retransmit. The sensor node will only establish neighborhoods for multimedia traffic delivery with other sensor nodes in the WSN that are using the same MIP, so all sensor nodes in the same cluster will have the same MIP. Head nodes exchange information and control messages with other head nodes. They maintain a database with existing head nodes and clusters and the MIPs associated with them. They will deliver multimedia data to other head nodes only if the destination cluster head node has the same MIP and will retransmit the multimedia flow to the cluster nodes if the multimedia flow belongs to the same MIP. A regular node can become a cluster head when the cluster head leaves the network or fails down.

The *Logic Management* level defines the protocol elements to manage the elements of the *Hardware Infrastructure*, by using the information received from the *Admin Interface* level. MIP is the logic element that gathers the information about the multimedia streams permitted in the cluster. It is the central element of the *Logic Management* level. In this level, the logical processes (discovery process, adjacency process, and forwarding process) that act over the sensor nodes as a function of their current state are also defined. When a sensor starts, it received the configured MIP from the *Admin Interface* level; then the discovery process is started and the node tries to find other nodes with the same MIP inside its coverage area. When it discovers other nodes, the adjacency process is started in order to create a neighborhood between both sensor nodes. These steps are followed by all new nodes in order to build the cluster. When a cluster is formed, it has the capacity to retransmit multimedia flows according to the ones defined in its MIP. Forwarding process is started when a sensor node creates a multimedia flow request or when a multimedia flow request is received from outside of the cluster (from other cluster or from outside the WSN through the gateway). It establishes the path to follow through the cluster and reserves the resources in every node belonging to the path. It makes possible the multimedia delivery and is responsible for guaranteeing the required QoS by the MIP during the communication.

Admin Interface level allows the interaction between the user and the sensor device. There is a graphic user interface (GUI) that lets the user modify the sensor init configuration, including the IP addressing and MIP selection. *Admin Interface* Level allows controlling manually the init process and disconnect process. The application also lets the user connect or disconnect the node to the WSN. The user can only make changes before the init process starts, so if a change is required, the sensor node must be stopped by using the disconnect process. Then, it should be initiated using the init process.

The number of available MIPs that can be selected by a sensor node, as well as the properties of each MIP, must be

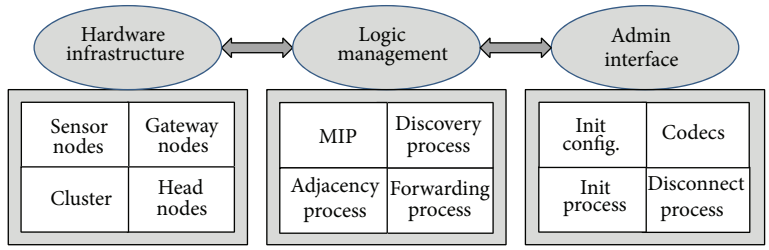


Figure 2: Elements of the architecture.

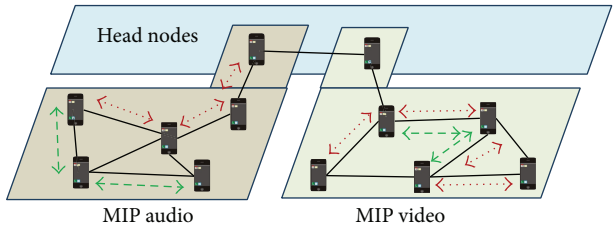


Figure 3: WSN structure based on MIP clusters.

Table 1: Protocol header.

1 byte		1 byte	0–255 bytes
2 bits	6 bits	8 bits	0–2040 bits
Version	Type	Length	Value

defined before the system is started. Each MIP represents a different type of multimedia traffic, so the MIP should be created taking into account the network characteristics, such as the nodes density, their location, node distribution, and radio coverage, jointly with the characteristics of the multimedia flow: type of traffic (audio or video), used codec, and QoS requirements. MIP definition is adapted to each particular case. For example, in a network topology with low nodes density and mixed video and audio flows, only two MIPS can be defined, one to create a cluster for audio delivery and another cluster for video delivery. But, if there is a network topology with high sensor nodes density, only dedicated to video delivery, but using a great variety of codec, several MIPs will be defined to split the multimedia flows that use video codecs in different clusters. The MIP assigned to the sensor includes the following information: maximum bandwidth (MaxBW) dedicated by the sensor node for retransmitting multimedia flows, minimum bandwidth (MinBW) required by a single multimedia flow to be processed, maximum delay (MaxDelay) permitted for the multimedia flow from the source to the destination, maximum jitter (MaxJitter) for a single multimedia flow and maximum hops (MaxHops) for a message in the WSN. Each MIP is identified by one-byte hexadecimal code, called HCode, and an alphanumeric code, called ACode.

Figure 3 shows the WSN MIP-based cluster structure. We have defined two MIPs: first one for audio flow delivery and the other for video flow delivery. Inside each cluster there could be simultaneous flow delivery with similar characteristics because they use the same MIP.

3.2. Protocol Fields. The developed protocol is included in the application layer of the TCP/IP stack protocols. UDP is chosen as encapsulation protocol at the transport layer in order to reduce the processing load of the sensor node, the bandwidth consumption, and the delay of the packets.

We wanted a simple protocol, with few fields, although it should be versatile. Protocol modifications should be easily made without big changes in the packet structure. Thus, we used the TLV (type-length-value) coding technique for the protocol implementation. TLV allows us to create new types of messages quickly and easily.

In Table 1, the protocol header fields are shown. We have included Version, Type, Length, and Value. The Type field allows us to interpret a received message. The information included in each type of message is variable and depends on the message objective, transmitter sensor node role, and receiver sensor node role. Generally, the size of the message is variable, so we have defined the Length field. It provides the length of the information carried at the Value field. Using TLV coding techniques increases flexibility and scalability of the protocol and these types of messages can be extended or be redefined in future revisions of the protocol very easily.

The protocol header fields are described below in greater detail.

(i) Version. This field provides the version of the protocol. Each version matches a specific and well-defined messages list. All devices in the WSN must use the same protocol version to communicate properly. The size of Version field is set to two bits in order to keep reduced to the size of the protocol message. The default value of the Version field is "00," which matches the protocol version 1.

(ii) Type. It is a numeric code used to identify the message type. Each message Type is defined in the specific protocol version. There is a message table which includes information about message length and how the message information carried at the Value field has to be interpreted on the reception

side. The size of the Type field is 6 bits, allowing a maximum of 64 message types.

(iii) Length. This field indicates the length of the Value field. The numeric value is given in bytes. The Value field is variable and its size depends on the type of message. The size of the Length field is one byte; the values range goes from 0 to 255. When the value of Length field is 0 it shows that the Value field does not exist; that is, it does not need to transmit any additional information.

(iv) Value. This field holds the information to be exchanged between the sensor nodes. The size of the field can take values between 0 and 255 bytes that matches the values of the Length field.

3.3. Data Structure. In order to carry out the required processes performed by the proposed protocol, the wireless sensor nodes have to exchange information. We have developed different types of messages with the purpose of performing next functions: exploring the network looking for devices with similar multimedia streaming purpose, creating sensor nodes adjacencies in order to build the cluster topology, sharing information about the sensor nodes status, their tables, and other network parameters, to start and run the multimedia flows through the cluster and notify to the neighbor nodes any event. Each defined message establishes the additional information to be included in the Value field. The following variables and structures were defined to facilitate the management of information.

(i) NODE_ID. It is the node identifier. This identifier must be unique across the whole network. The NODE_ID parameter size is 2 bytes and its value should be set before the init process starts, at the initialization process. There are 3 different mechanisms to generate a sensor NODE_ID: (1) static configuration, the identifier is manually defined by the sensor administrator; (2) automatic configuration, the last two bytes of the IP address are used as NODE_ID; and (3) dynamic configuration, where a network service uses the Multicast IP address 239.100.100.255. This last configuration option requires the previous configuration of one or more nodes as servers with preconfigured tables in order to assign the NODE_ID. This option has been only designed for testing and to facilitate the research work, but it is discouraged to use it in real environments because it introduces the need of servers.

(ii) NODE_RESOURCES. This variable contains information about the available bandwidth of the sensor node for multimedia delivery. The size of this variable is two bytes. The bandwidth is measured in Kbps. The initial value of the variable is set to the MaxBW value of the assigned MIP. When a new resource reservation for multimedia delivery is made, the NODE_RESOURCES value is decremented until the resource reservation is canceled or the delivery ends. When the NODE_RESOURCES value is below the MinBW parameter then the node changes its value to zero and no new multimedia delivery is allowed.

(iii) NODE_ADJ. It is a data structure representing the connectivity state of a node into the cluster. The size of this parameter is variable and ranges between 1 and 511 bytes. First byte shows the number of adjacencies of the node in that moment. Then, the data structure is built by concatenating the NODE_ID of the neighbor node who has established a successful adjacency with. When the node starts and it has not still been established any adjacency, the initial value of NODE_ADJ is set to 0x00 and is 1 byte in size. When the first adjacency is created the first byte is changed to 0x01 and the neighbor NODE_ID value is joined. From this point, every time a new adjacency is created, the first byte will be incremented and the new NODE_ID value will be concatenated to the NODE_ADJ structure. Because of data structure limitations, the maximum number of adjacencies by a node is limited to 255 adjacencies.

(iv) NODE_NCON. It indicates the total local number of properly established and active adjacencies. The size of the variable is 1 byte. Its initial value is set to 0x00. This variable matches the value of the first byte on the NODE_ADJ data structure. NCON value is incremented or decremented each time an adjacency is created or destroyed.

(v) NODE_NSEQ. This variable represents the version number of the state table of the sensor device. The parameter size is 2 bytes. When the sensor node starts, the initialization process set its value to 0x0000. When a state change occurs, for example, when an adjacency with other node of the WSN is created or destroyed, the NODE_NSEQ value is increased or decreased. Then, the system sends a cluster state update (CSU) message to all nodes with successful adjacencies to update their state table. When a CSU message is received, the node compares the NODE_ID and NODE_NSEQ values on the received message with the information stored in its state table. If the value of NODE_NSEQ for this NODE_ID in the state table is below the received value, the state table is updated with the information included in the CSU message. Then, the message is forwarded to all local adjacencies except to the neighbor that sent the original CSU message. If NODE_NSEQ of the CSU message is equal to or lower than the values of the state table, the CSU message is discarded.

(vi) NODE_STATE. It is a data structure created by concatenating other local variables and structures: NODE_NSEQ, NODE_ID, NODE_RESOURCES, and NODE_ADJ variables. The data structure size is calculated as a function of the number of adjacencies established by NODE_NCON value, and it ranges between 7 bytes, when there is not created any adjacency, and 517 bytes, when the maximum value of adjacencies has been reached.

(vii) CSU_NSEQ. This variable is a sequence number value used in CSU messages in order to allow message fragmentation. When the size of the information in the state table cannot be fit into a single message, the CSU sequence number

allows fragmenting the information into multiple messages sequentially numbered. The field size is 1 byte and the default value is set to 0x80 when no fragmentation is needed. When fragmentation is used, the packets are consecutively numbered starting from 0x01. Possible values range between 0x01 and 0xEF. When the last fragment is sent, the sequence number is increased from the previous message and, then, the first bit is changed to "1" indicating that this is the last fragment of the sequence.

(viii) CLUSTER_MIP. The CLUSTER_MIP value matches the HCode value of the assigned MIP. This is a 1-byte variable. This parameter is exchanged between neighbor nodes in the adjacency process. The MIP table is defined for the whole WSN. The number and characteristics of the defined MIP depend on the traffic pattern and multimedia flows of the network.

(ix) CLUSTER_N. This parameter is used to distinguish two different clusters using the same MIP. The size of the variable is 1 byte. When the first cluster node creates the cluster and it is not aware of other clusters with the same MIP, then it selects the CLUSTER_N value equal to 0x00. Next, the first cluster node sends a request to existing cluster heads in order to know their CLUSTER_N. After this step it becomes cluster head and adds next available value from the received replies to its CLUSTER_N parameter.

(x) CLUSTER_ID. It is the cluster identifier. This value must be unique for each cluster within the same WSN. Two independent clusters into the WSN can share the same MIP but they must always have different CLUSTER_ID value. The size of the variable is two bytes. Its value is established by the first node in the cluster. The first node is defined as the node that receives the discovery message ACK to establish the first cluster adjacency. The CLUSTER_ID value is built by concatenating two variables, CLUSTER_MIP and CLUSTER_N. In case of CLUSTER_ID duplications in the same WSN (because of lost messages or formed cluster joining), the oldest cluster keeps its CLUSTER_ID, and the youngest cluster changes its value to the next free value. An update message is sent to all nodes into the cluster to notify and update the new CLUSTER_ID.

(xi) CLUSTER_DIAMETER. This variable shows the current cluster diameter. The cluster diameter is defined by the highest value of the lowest distance between any two nodes in the cluster. Distance between two sensor nodes is calculated by the routing algorithm. It is measured in number of hops. The size of the variable is 1 byte. When a sensor node starts, it has not established any adjacency yet, and then the CLUSTER_DIAMETER value is set to 0. Later, when the first adjacency in the cluster is created, the value is changed to 1 on both nodes. Each time a new sensor node is added to the cluster topology, the CLUSTER_DIAMETER is recalculated using the routing protocol in order to guarantee that it does not overcome the Maxhops value established in the MIP cluster. If the new adjacency exceeds Maxhops, then adjacency fails to be established.

(xii) MEDIA_RESOURCES. This parameter identifies the bandwidth resources needed for a single multimedia communication. This variable is used when the sensor node creates and processes a new delivery request. Possible values can vary from MinBW to MaxBW of the assigned MIP. It depends on the characteristics of the codec used for multimedia delivery. Its value represents the bandwidth measured in Kbps and it is 2 bytes long.

(xiii) MEDIA_SOURCE. When a request for resource reservation takes place, the NODE_ID value of the source node (SN) is copied in this variable. SN is the sensor node where the multimedia delivery was originated inside the cluster. Like NODE_ID variable, the MEDIA_SOURCE size is 2 bytes. The origin of the multimedia delivery can be located outside the WSN; in this case the SN is defined as the gateway node used to enter the WSN.

(xiv) MEDIA_TARGET. This variable carries the NODE_ID value of the target node (TN). In a similar way as the SN was defined, the TN is the sensor node where the multimedia transmission ends inside the cluster. Its size is also 2 bytes. As in the previous case, the multimedia communication may finish outside the WSN, through a gateway sensor node connected to an external network. In this case, the MEDIA_TARGET is defined as the NODE_ID of the gateway node.

(xv) MEDIA_ROUTE. This structure contains the full route for a multimedia packet flowing from the MEDIA_SOURCE to the MEDIA_TARGET. It is built by adding every sensor NODE_ID on the route. The route is calculated by the routing algorithm. Its size can vary from 4 bytes, when SN and TN have established a valid adjacency, to 32 bytes, when there are 16 hops on the route, the maximum number of allowed hops for any cluster. The first NODE_ID used to build the structure is the MEDIA_SOURCE and the last matches the MEDIA_TARGET.

(xvi) MEDIA_NHOP. This variable has the number of hops between MEDIA_TARGET MEDIA_SOURCE as it is calculated from the routing algorithm in the MEDIA_SOURCE sensor node. The size of this parameter is 1 byte. The maximum number of hops allowed by the protocol implementation inside a single cluster of the WSN is set to 16 hops. However, the number of hops between any two nodes on a specific cluster can never be greater than the CLUSTER_DIAMETER parameter (as it is defined in the assigned MIP).

(xvii) MEDIA_NSEQ. This is the sequence number assigned to a multimedia delivery for the source node. The size of the variable is 2 bytes. The initial value is set to the hexadecimal value 0x0000. Each time a new request for multimedia delivery is originated in a sensor node the MEDIA_NSEQ value is incremented by one. This variable allows the protocol to differentiate between several multimedia flows being delivered simultaneously from the same source node.

(xviii) MEDIA_INFO. This is a data structure that contains the whole information for a single multimedia delivery that is needed and used by the remaining cluster sensor nodes. It is built on the SN when a new multimedia request is originated. The following parameters and structures are added in order to build the MEDIA_INFO structure: MEDIA_RESOURCES + MEDIA_NSEQ + MEDIA_NHOP + MEDIA_ROUTE. The size of the data structure depends on the number of hops on the route indicated by the MEDIA_NHOP. The size can vary between 11 and 39 bytes.

3.4. Message Table. The messages used by the protocol are described in this section. Here we define the version 1 of the protocol. The TLV coding used by the protocol encapsulation allows us to change the list of messages in the following versions. For a better understanding, the whole list of messages has been organized considering the system process they belong to. UDP protocol is selected at the transport layer. Despite this, relevant messages need to be confirmed. For example "ACK Discovery" is a confirmation message for the "Discovery" message and "Confirm Join" message confirms the "Request Join."

Table 2 shows the protocol messages used at the adjacency process. Messages belonging to the adjacency process are shown on Table 3. Table 4 describes the forwarding process messages and the disconnect process messages are listed on Table 5. System processes are detailed in the next section.

4. System Operation

This section details the protocol operation. There are four main processes: discovery, adjacency, forwarding, and disconnect.

4.1. Discovery Process. Figure 4 shows the messages exchanged in the discovery process. A sensor node starts the discovery process when the sensor node initialization process has finished. This sensor node is named new node (NN). The NN changes to the discovering state and it begins sending messages looking for other sensor nodes. A "Discovery" message is sent every 60 seconds. If there are not answers after three messages, the sensor node stops sending "Discovery" messages. The Value field at the "Discovery" message has the NODE_ID, to identify the CN, and the CLUSTER_MIP to inform the selected MIP. Messages are sent to the Multicast IP address "239.100.100.CLUSTER_MIP," where the last byte matches the CLUSTER_MIP parameter. Thus only sensor nodes with the same MIP and listening to the Multicast IP address, will receive the messages. The receiver sensor nodes are called border cluster node (BCN). BCN replies by sending the "ACK Discovery" message to the NN. The "ACK Discovery" messages are sent to the unicast IP address of the new sensor node and the multicast address is not used anymore. The "ACK Discovery" message has the following information: The BCN NODE_ID, the NODE_NCON that shows the amount of established adjacencies, the CLUSTER_ID to identify the cluster, and the CLUSTER_DIAMETER. When the NN receives the "ACK Discovery" message, it compares the CLUSTER_DIAMETER with the MAX_HOPS parameter; if both values are equal, then the adjacency process finishes here. If two of more clusters are available, the NODE_NCON information is used by the new sensor node in order to select the most appropriate cluster to connect with. The lowest value is preferred.

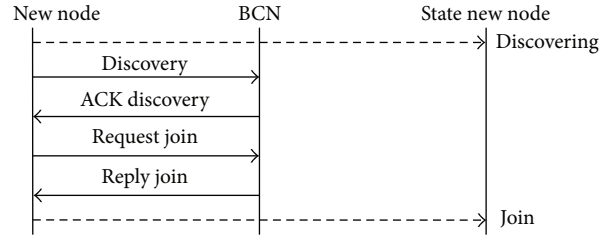


Figure 4: Discovery process.

The new sensor node keeps waiting at least for 60 seconds after sending the first "Discovery" message and before selecting the target cluster; thus it allows arriving on time other possible "ACK Discovery" messages from different BCNs. When a valid cluster is discovered, and it is selected, the candidate sensor node sends a "Request Join" message to the selected sensor node (or nodes if they belong to the same selected cluster). This message contains the BCN NODE_ID that it is looking to build the adjacency, the CLUSTER_ID it wants to join and the available resources in the candidate sensor node through the NODE_RESOURCES parameter. The BCN uses the information about the NN resources to update its own state table and to notify the other sensor nodes in the cluster topology. Then, it replies the NN by sending a "Reply Join" message; it changes to the Join state and the discovery process ends. If two or more BCNs from the same cluster are discovered the sensor node will have adjacencies with all of them.

4.2. Adjacency Process. The state table holds the information about all sensor nodes belonging to the same cluster. All sensor nodes in the same cluster share the same state table. There is a table entry for each sensor node in the cluster; thus when a new adjacency appears, the full state table is exchanged. Each table entry is stored in a single NODE_STATE structure. This structure keeps the following information about a single sensor node: NODE_ID, available resources, number of adjacencies, neighbors NODE_ID, and NODE_NSEQ. If it is the first adjacency of the new sensor node, then there is only one entry on its state table and this is about its own link-status information.

The Adjacency process begins when the new sensor node makes a transition to the Join state. The exchange of messages in the adjacency process is displayed in Figure 5. This image represents the specific case when the adjacency between two sensor nodes, a NN and a BCN, is successfully completed. The inside cluster node (ICN) is defined as any other sensor node inside the cluster that is not going to build a direct adjacency with the new sensor node. A "Cluster State Update (CSU)" message is sent from the NN to

Table 2: Discovery process messages.

Discovery process				
Message	Type	Length	Value	Description
Discovery	0x01	3 Bytes	NODE_ID CLUSTER_MIP	Message looking for neighbors to join or to create a cluster
ACK discovery	0x02	6 Bytes	NODE_ID NODE_NCON CLUSTER_ID CLUSTER_DIAMETER	Confirms the CLUSTER_ID available to join
Request join	0x03	6 Bytes	NODE_ID NODE_RESOURCES CLUSTER_ID	Sensor node sends a request to build a new adjacency
Confirm join	0x04	4 Bytes	NODE_ID CLUSTER_ID	Sensor node confirms the request to join

Table 3: Adjacency process messages.

Adjacency process				
Message	Type	Length	Value	Description
Cluster state update (CSU)	0x11	10–255 bytes	CSU_NSEQ NODE_ID NODE_STATE NODE_STATE ... NODE_STATE	To exchange the state table information between adjacency nodes. When state table information exceeds the 255-byte limit; it is fragmented in two or more messages.
ACK cluster state update (ACK CSU)	0x12	3 bytes	CSU_NSEQ NODE_ID	Confirmation message of a CSU message
Node state update (NSU)	0x13	8–255 bytes	CSU_NSEQ NODE_STATE	Information about the state of a single node. If the NODE_STATE variable exceeds 255 bytes information is fragmented in two or more messages.
ACK node state update (ACK NSU)	0x14	3 bytes	CSU_NSEQ NODE_ID	Confirmation message of a NSU message
Cluster join	0x15	4 bytes	NODE_ID CLUSTER_ID	The node has filled its state table with the cluster information and it requests to join the cluster
ACK cluster join	0x16	4 bytes	NODE_ID CLUSTER_ID	Confirmation message of a cluster join message

Table 4: Forwarding process messages.

Forwarding process				
Message	Type	Length	Value	Description
Request forwarding	0x21	13–41 bytes	NODE_ID MEDIA_INFO	A new multimedia flow requests a resource reservation. This message is sent from the source node to the target node
Reserve resources	0x22	13–41 bytes	NODE_ID MEDIA_INFO	Target node sends a resource confirmation to the source node
Confirm reserve resources	0x23	13–41 bytes	NODE_ID MEDIA_INFO	Reservation confirmation from the source node to the target node
Queue reserve	0x24	13–41 bytes	NODE_ID MEDIA_INFO	Notification message when multimedia flow is placed in queue
Reject reserve	0x25	13–41 bytes	NODE_ID MEDIA_INFO	Reservation cancelation
End transmission	0x26	13–41 bytes	NODE_ID MEDIA_INFO	Multimedia delivery is finished and resources have to be released

Table 5: Disconnect process messages.

Disconnect process				
Message	Type	Length	Value	Description
Disconnect	0x31	2 Bytes	NODE_ID	A node is breaking an adjacency
ACK disconnect	0x32	2 Bytes	NODE_ID	Confirmation of the disconnect message

the BCN. The message is built with the NN NODE_ID and the CSU_NSEQ. The sequence number is used to fragment the "CSU" message information when necessary. Moreover, full information on the NN state table is included in the message; the NODE_STATE structure is used here. The "CSU" message needs always to be acknowledged by an "ACK CSU" message from the BCN; if any "ACK CSU" message is not received in 10 seconds, after the "CSU" message was sent, it is sent it again. If a sensor node sends the same "CSU" message three times, and it does not receive any answer, the adjacency process finishes unsuccessfully. After the "ACK CSU" message, the BCN sends its own state table information to the NN by sending one or more "CSU" messages. The state table is encoded in NODE_STATE structures as table entries. If the message size exceeds the limit of 255 bytes, then the information is fragmented to be sent on several "CSU" messages. The CSU_NSEQ value is used to allow fragmentation. Each "CSU" message needs to be acknowledged by an individual "CSU ACK" message in order to avoid losing information; neighbor sensor nodes need to keep the same state table; otherwise routing algorithm will not be able to calculate the most appropriated route between sensor nodes in the same cluster. After the state table has been fully exchanged between the NN and the BCN, the NN makes a transition to the associated state. This is a transitional state, both sensor nodes are sharing the whole cluster link-state information but they have not completed their adjacency yet. BCN has not updated any other ICN yet.

When the NN changes to the associated state it sends a "Cluster Join" message to the BCN. "Cluster Join" message has the CLUSTER_ID it is trying to join. The BCN updates its state table with the STATE_NODE information of the NN and it increases by one its NODE_NSEQ value. Then, the BCN sends a "Cluster Join ACK" message to the NN in order to accept the new adjacency. At this moment, the NN makes a transition to the established state, which means that it has joined the cluster. The adjacency process with the NN is completed. Finally, the NN information is flooded to the rest of the sensor nodes in the cluster by sending two "NSU" messages to all sensor nodes in the cluster. The main difference between "CSU" and "NSU" messages is that "CSU" message carries the full state table information, but the "NSU" message only carries an individual table entry for a single sensor node. In this case, two "NSU" messages should to be sent: one for the NN information and the other for the BCN updated information. Every ICN checks the NODE_NSEQ for each message; then, it updates its state table and, finally, forwards the "NSU" message to all its neighbors, except to the one it has received the message from. If the NODE_NSEQ in the received STATE_NODE structure is equal or greater than the NODE_NSEQ in the ICN state table "NSU" message is ignored. Each "NSU" message needs to be acknowledged by an "ACK NSU" message, even when the "NSU" message is ignored.

4.3. Forwarding Process. The forwarding process starts when there is a request for multimedia delivery in the cluster. Figure 6 shows the message flow diagram for the forwarding process. The example detailed in the figure explains how a new multimedia delivery is requesting a resource reservation. The request is queued by a starved node without enough resources and finally it is processed when resources are released at the queued sensor node. Source node (SN) is defined as the first sensor node in the cluster where the multimedia request takes places. SN can be a gateway node, if the request is generated outside the WSN, or it can be any other sensor node if the request is generated inside the WSN. Target node (TN) is the destination multimedia flow inside the WSN; it can be a gateway node if the IP address destination is outside the WSN. In the diagram, the first hop node (FHN) has been defined as the first cluster node on the path to the target node. FHN is calculated by the routing algorithm starting from the SN neighbors. In this case, ICN will be those nodes on the path between the SN and the TN.

When the forwarding process starts, the SN is in the established state or in the forwarding state and it receives a new multimedia flow request. First, it checks if there are enough local resources to process it. If the SN has enough available resources, the full path to the TN is calculated. The routing algorithm is used only once and only at the SN; the SN state table information contains the whole information about the cluster needed to establish each hop on the path; thus the path cannot be modified.

The message exchange starts when a SN sends a "Request Forwarding" message to the first hop node (FHN), which is the first NODE_ID on the calculated path. The message holds the SN NODE_ID and the MEDIA_INFO data structure. The MEDIA_INFO structure provides complete information about the multimedia request: MEDIA_RESOURCES, MEDIA_NSEQ, MEDIA_NHOP, and MEDIA_ROUTE. The ROUTE_MEDIA structure contains the NODE_ID from all hops on the path, from the SN to the TN. MEDIA_RESOURCES show the bandwidth resources required to enable to process the multimedia communication. MEDIA_NHOP matches the amount of hops on the path. MEDIA_NSEQ is the sequence number assigned by the SN to identify this particular multimedia flow.

The FHN receives the "Request Forwarding" message and checks if its NODE_ID is included in the MEDIA_ROUTE structure. If not, the message is discarded. If it is inside, the FHN checks it resources availability and the value is compared to the MEDIA_RESOURCES value. If there are enough local resources, the FHN reads the next NODE_ID on the hop list and the "Request Forwarding" message is forwarded to

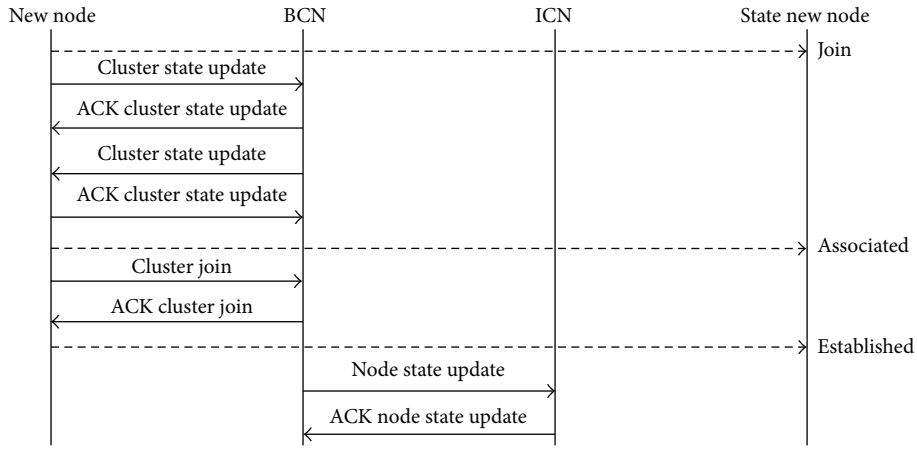


Figure 5: Adjacency process.

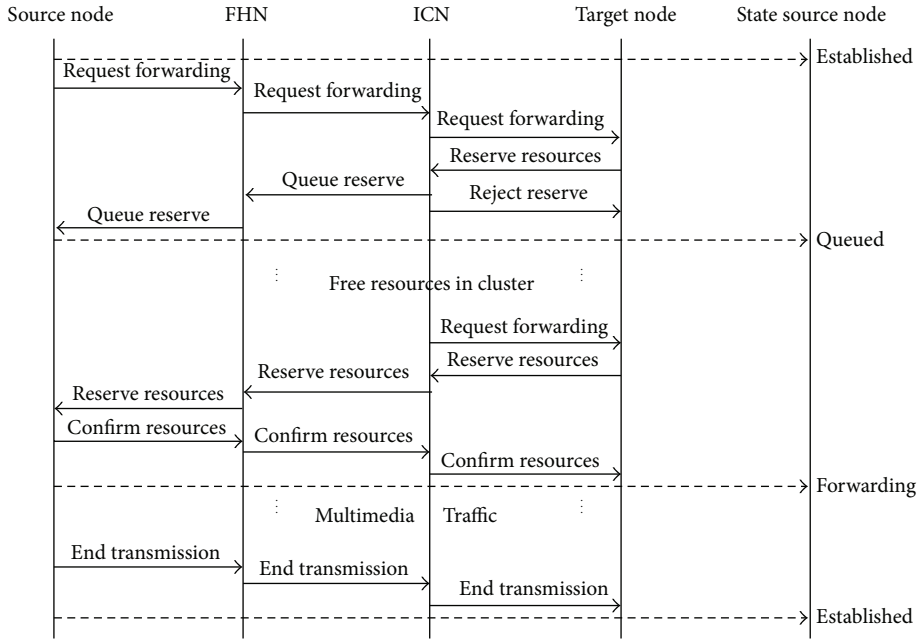


Figure 6: Forwarding process.

it. Resource reservation at the FHN is not set yet and can be used by current traffic, but the bandwidth resources of this request will not be used by other request reservation until the reservation is confirmed or rejected. All hops on the path perform the same process, hop by hop, in order to reach the TN. Finally, the TN receives the "Forwarding Request" message. The TN NODE_ID is compared with the last hop on the list provided by the MEDIA_ROUTE structure in order to check that the TN is included in this multimedia request. Available resources are checked as the other sensor nodes on the path. If there are enough resources, a reservation is made. The NODE_RESOURCES variable is decremented in the amount indicated by the MEDIA_RESOURCES parameter. This is a temporary reservation and it needs to be confirmed by the SN. Thus, the TN sends a "Reserve Resources" message back to the SN. This message also carries the MEDIA_INFO structure and it should follow the same path of the "Forwarding Request" message, but in the opposite direction. Each sensor node on the path performs a temporary reservation and follows the message back to reach the SN.

If a sensor node on the path cannot make the reservation because there is not enough bandwidth available to guarantee the multimedia communication, the designed protocol can put the request in queue for this sensor node. This process is shown in Figure 6, where the ICN decreases its bandwidth when it receives the "Reserve Resources" message. When an ICN is congested it can perform three different actions: it stores the request in a waiting queue, then it sends a "Reject Resources" message to the TN and finally, and it sends a "Queue Reserve" message to the SN. Both messages use

the MEDIA_ROUTE information in order to repeat the same path and to inform all sensor nodes in the path. The temporary resources reservation made in the sensor nodes between the ICN and the TN are cancelled by the "Reject Resources" message. On the other side, the "Queue Reserve" message releases the prereservation made at the sensor nodes between the SN and the ICN. The SN puts the multimedia request in a request queue and it waits to receive a notification from the congested sensor node when requested resources will be available. Both, "Reject Resources" and "Queue Reserve" messages have the NODE_ID field with the NODE_ID of the congested sensor node, ICN; thus all sensor nodes on the path can locate the congestion problems. Routing algorithm will not include congested nodes in the path. There are two options when the SN receives the "Queued Reserve" message: (1) it can wait for released resources in congested nodes or (2) it can calculate again the path to the TN but avoiding the congested sensor node. In this second case, the SN sends a "Reject Reserve" message to the congested sensor node; the waiting queue in the congested sensor node will be deleted. If the SN keeps the request in queue, then a timer is started in order to prevent a blocked multimedia communication. When the timer expires, a "Reject Reserve" message is sent to the congested sensor node.

If the congested sensor node keeps the request queued, it waits till other multimedia communications ends in order to have enough bandwidth resources. The forwarding process is started again from this point. Figure 6 shows the exchanged messages. Congested sensor node sends a "Request Forwarding" message to the TN. The original MEDIA_INFO is used. Then, a "Reserve Resources" message is sent back to the SN again from the TN. Finally, since all sensor nodes in this example have enough available resources to make the reservation, the "Reserve Resources" message reaches the SN. SN knows that all sensor nodes on the path to the TN have enough resources and they have made a temporary reservation to process the request. Then, SN sends a "Confirm Resources" message to the TN through the MEDIA_ROUTE and temporary reservations are confirmed. The SN changes to the forwarding state and the multimedia delivery begins.

When the multimedia delivery ends, SN sends an "End Transmission" message. This message carries the MEDIA_INFO structure, which is sent to the TN to inform each sensor node that the delivery has finished and the allocated resources can be released.

4.4. Disconnect Process. Figure 7 shows the exchanged messages in the disconnect process. Disconnect process is started by the sensor node to shut down or reboot. The sensor node sends a "Disconnect" message to each neighbor. Then, a 10-second timer is activated waiting the "ACK Disconnect" message. If no neighbor sends the "ACK Disconnect" message in the timer interval, then the "Disconnect" message is sent again until 3 times. After it, the sensor node leaves the cluster.

Next, neighbor sensor nodes update their status table. All information about the disconnected sensor node is removed. Then, they send a "NSU" message to their neighbors. The "NSU" message is flooded across the cluster, in order to let all sensor nodes update their state table. Every "NSU" message is acknowledged by the "ACK NSU" message.

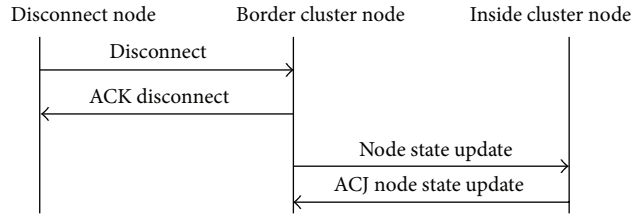


Figure 7: Disconnect process.

5. Performance Study

In order to validate the proposed algorithm we have designed and built a test bench. Our protocol organizes sensor nodes in four clusters: two audio clusters and two video clusters. Each cluster has assigned a different MIP. When a NN starts in the wireless network it knows the MIP that it belongs to. Then it tries to discover other sensor nodes with the same MIP and finally it joins the cluster. If it is the first sensor node in the network with this particular MIP the sensor node keeps waiting for new sensor nodes with the same MIP.

Several topologies arrangements have been studied in order to know how the quality of service parameters change when the diameter of the topology increases. The QoS parameters, delay, jitter, and packet loss have been measured for each MIP cluster in three experimental conditions: cluster diameter of one hop, two hops, and three hops.

Before the wireless sensors start, they have been configured with static IP address and wireless ad hoc network configuration, wireless channel, and interface speed. IEEE 802.11g standard has been selected as the wireless technology for the wireless sensor nodes.

The four MIPs are simultaneously working in the same WSN. Two audio MIPs have been selected: AUDIO_64K and AUDIO_192K. First, the AUDIO_64K matches the regular audio communications and audio IP calls performed through the PCM codification standard and the G.711 codec, the most compatible and widely used at all kind of audio applications and protocols. These deliveries offer a sound quality similar to the quality of a phone line. The AUDIO_192K MIP matches codecs used at high quality audio communications. With this kind of codecs it is possible to deliver music and human voice with nearly perfect quality.

For video deliveries we have chosen two MIPs: VIDEO_1500K and VIDEO_3500K. The first MIP, VIDEO_1500K, has been chosen because it represents the quality for a video delivery performed in high definition TV (HDTV) with 720p format. In the same way, the VIDEO_3500K is included because it is a typical standard delivery for 1080p format in HDTV.

5.1. MIP Comparison. The first test bench was set to find out if there are differences between multimedia deliveries belonging to different MIPs when they take place over similar

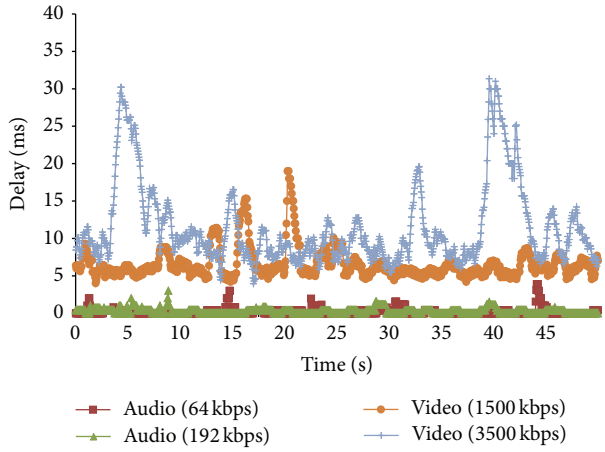


Figure 8: Delay as a function of the time for each MIP.

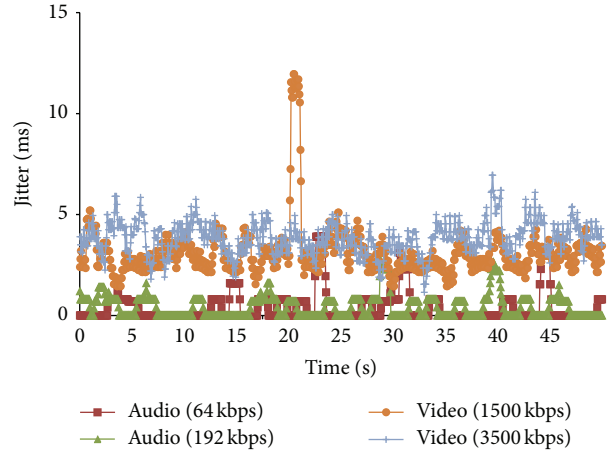


Figure 9: Jitter as a function of time for each MIP studied.

cluster topologies. In order to study the cluster behaviour in terms of QoS parameters for each MIP, the WSN topology was designed with the aim of building four clusters, one cluster for each MIP. In this experimental design the maximum number of hops for every cluster was established at two hops. In order to be able to compare the obtained results for each cluster, several environment variables and experimental conditions have been controlled: there are the same number of sensor nodes at each cluster, the same average distance between sensor nodes in each cluster, only one multimedia delivery is in progress at a time, and the noise level at the 2.4 GHz microwave band is measured and controlled. Figure 8 shows the delay measured for four different MIPs through clusters with identical characteristics but with different MIP settings. The figure represents the average delay of the last 20 samples received at any time. In order to estimate the average delay, we applied (1) on the obtained measurements:

$$Y_i = \frac{\sum_{j=i}^{i+20} X_j}{20}. \quad (1)$$

Both studied audio codecs have obtained similar delay results. Figure 8 shows that the average delay remains below 5 milliseconds when audio is delivered. These results indicate that the quality of audio transmission can be performed over this cluster without any loss of quality, even high quality audio with 192 Kbps. Results for video codecs seem not to be as good as for audio codecs. However, the average delay for video delivery is always below 30 milliseconds (there are two peaks of about 30 milliseconds at the 5th second and at the 40th second) and these values are enough to guarantee an excellent quality on video regular communications.

For the same number of hops, we observe that the delay is rising when the bandwidth spent for multimedia communication through the cluster grows. The behaviour of audio codecs compared to video codecs is clearly different. In order to determine if there is a significant difference between both audio codecs and between both video codecs, we need to make the statistical analysis of the experimental data. Table 6 shows the analysis results. The 99% confidence interval was calculated for an average delay of each experimental condition (α = 0,01). In order to establish relationship between each series, three null hypothesis were assumed: there are not differences between audio and video measures, there are not differences between two audio measures with different bandwidth consumption, and there are not differences between two audio measures with different bandwidth consumption

As expected, mean delay values are significantly different when any audio codec is compared with any video codec, so we can completely reject the null hypothesis and accept the alternative hypothesis: difference between delay of audio and video MIPs has statistical significance. In the same way, when the mean delays for both video MIPS are compared a statistical significant difference can be concluded. However, when audio MIPs with different bandwidth consumption, 64 Kbps and 192 Kbps, are compared, it is not possible to deduce any significant difference because the mean delay of one audio MIP is inside the confidence interval of the other MIP. In the last case, it is not possible to reject the null hypothesis at least with $P = 0.01$.

Figure 9 shows the jitter obtained in the experimental tests. As it happens with the delay results, we can see that the jitter for audio cluster is significantly lower than the jitter for video delivery. The second important result is that all multimedia delivery has jitter values below 15 milliseconds. Only few samples are over the 10 milliseconds. The quality of a multimedia communication can be affected by jitter values when they are as low as 20 or 30 milliseconds, but it is possible to easily manage a jitter value of 15 milliseconds building a buffer in the receiver side to eliminate its harmful effect.

Data was analyzed to know if there are some significant differences between two similar MIPs, that is, two audio MIPs or two video MIPs. Table 7 shows the statistical parameters for each data series with $\alpha = 0.01$. Statistical inference has been conducted as the previous delay analysis. Mean jitter values for AUDIO_64K and AUDIO_192K are very similar. Even the AUDIO_192K shows a mean jitter a bit bigger than AUDIO_64K. However, mean value of the first data series is included at the confidence interval of the second and vice versa. Null hypothesis cannot be rejected and it is not possible

Table 6: Statistical values and confidence interval for delay as a function of the MIP.

MIP	Parameters						
	N	μ (ms)	σ (ms)	Min (ms)	Max (ms)	Confidence interval (ms)	
AUDIO_64K	1000	0.276	0.696	1	4	0.204	0.348
AUDIO_192K	1000	0.205	0.282	1	4	0.175	0.234
VIDEO_1500K	1000	6.588	2.089	4	19	6.370	6.805
VIDEO_3500K	1000	11.461	5.509	5	31	10.887	12.034

to deduce any difference between both audio data series. By contrast, differences between mean jitter for both video MIP can be accepted. Null hypothesis is rejected in this case. Moreover, the null hypothesis is rejected between the mean jitter values of audio and video MIPs.

Based on these results, we can conclude that, in this experimental setup, delay and jitter parameters obtained using different video MIPs are different. Moreover, obtained QoS parameters of video MIPs are different than the QoS parameters of audio MIPs. These results confirm the benefits to divide the whole WSN into several clusters based on MIP configuration. Clusters with multimedia video traffic have a different QoS behaviour as a function of the features of the video delivery and they are also different than the audio delivery. Keeping separate multimedia flows through the MIP architecture allows the network to improve the delay and jitter parameters for multimedia delivery with low requirements.

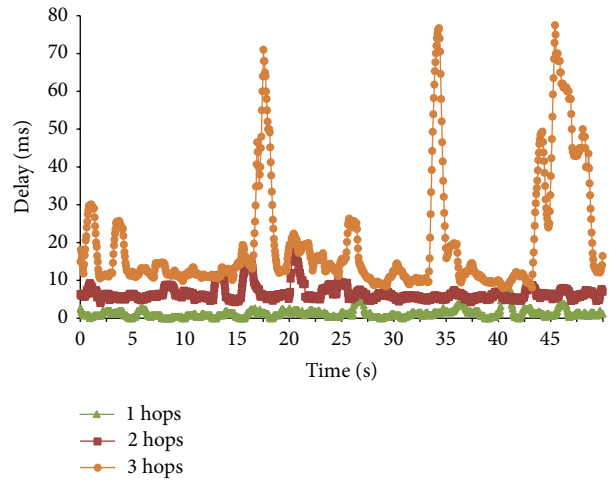


Figure 10: Delay as a function of time for different number of hops.

5.2. Cluster Comparison. In this second experiment we have studied the number of hops in the cluster. In order to perform this study only one MIP was selected VIDEO_1500K. The WSN topology was modified to achieve three different cluster diameters. Multimedia delivery was always performed through the maximum number of hops allowed in each cluster topology. The number of hops selected for the three experiments were: one, two, and three hops.

Figure 10 shows the results obtained for the delay as a function of the time in the three cases. The main result was that the delay is worse when the number of hops increases, but three-hop case has very high peaks, which might be taken into account. Delay for 1 hop delivery is minimal; values were only few milliseconds above zero, and there was not any big value in the whole series; all values were below 100 milliseconds. Delay for 2-hop condition moves between 5 and 10 milliseconds; there were some peaks on the graph; however they are of small size. Finally, delay for 3-hop transmission was the biggest with values between 10 and 20 milliseconds; there are also many peaks with mean values above 70 milliseconds. Multimedia delivery can be optimal with values of up to 150 milliseconds; above this limit quality of service would be decreased. However, it should be noted that the measured delay is only the delay introduced by the sensor nodes transmission on the cluster, but in a real case there are other processes and transmissions out of the cluster that need to be considered to calculate the final delay.

In order to corroborate the correct interpretation of these results, a statistical analysis was performed. Table 8 shows the statistical analysis. In the inference analysis, $\alpha = 0.01$ is assumed and the 99% confident interval was calculated for each experimental condition. Two null hypotheses have been stated: there are no differences between the mean delays in the cluster with one hop and the cluster with two hops, and there are no differences between mean delay in the cluster with one hop and the cluster with two hops.

Mean delay of each series is outside the confidence interval of the remaining cases. Both null hypotheses can be fully rejected with 99% probability. It is possible to affirm that the mean delay value through a two-hop cluster is bigger than through a one-hop cluster, and the mean delay value through a three-hop cluster is bigger than through a two-hop cluster.

Jitter values are shown in Figure 11. We can see that the values for data series of 1 and 2 hops are very similar, with an average below 5 milliseconds. Otherwise, the 3-hop series shows higher values, around 10 milliseconds, but it is always below 15 milliseconds. These jitter results have been obtained through the control of some experimental conditions: there was only one delivery, reduced noise level, and so on. But in a real environment there are a lot of variables that can affect the multimedia delivery. Thus, a 10 milliseconds level of jitter obtained in these ideal conditions should be interpreted with caution.

Jitter measures for one and two hops are very similar and we need to make the statistical inference analysis to know the relationship between these data series. The analysis is conducted following the same criteria than the previous delay analysis. It is shown in Table 9.

Table 7: Statistical values and confidence interval for Jitter as a function of the MIP.

MIP	Parameters						
	N	μ (ms)	σ (ms)	Min (ms)	Max (ms)	Confidence interval (ms)	
AUDIO_64K	1000	0.412	0.784	0	4	0.348	0.475
AUDIO_192K	1000	0.410	0.556	0	3	0.365	0.455
VIDEO_1500K	1000	3.104	1.315	1	12	2.996	3.211
VIDEO_3500K	1000	3.824	0.786	1	7	3.760	3.888

Table 8: Statistical values and confidence interval for delay as a function of the number of hops.

Hops	Parameters						
	N	μ (ms)	σ (ms)	Min (ms)	Max (ms)	Confidence interval (ms)	
1	1000	1.374	1.269	0	8	1.271	1.477
2	1000	6.589	2.089	4	19	6.419	6.759
3	1000	19.797	15.119	8	78	18.565	21.028

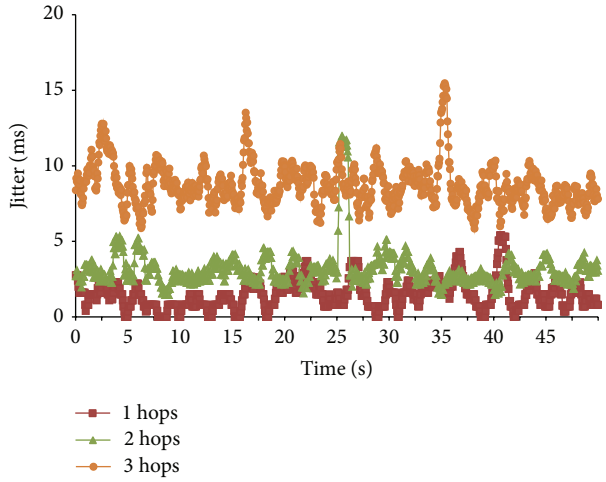


Figure 11: Jitter as a function of time for different number of hops.

As it happens in the case of delay, null hypothesis can be rejected and the alternative hypothesis is accepted. Mean jitter values for 2 hops series data is significantly bigger than for 1 hop, and mean jitter value for 3 hops is bigger than for 2 hops.

From these results we can conclude that, in this experimental environment, the cluster diameter may negatively affect the multimedia traffic QoS parameters. The proposal architecture and the developed network protocol can improve QoS parameters by minimizing the maximum number of hops into the cluster.

5.3. Packet Loss Study. Table 10 shows packet loss percentages for each experimental case. We can see that there is very few percentage of lost packets when there is only one hop in the WSN. When the topology becomes more complex, and packets have to make two hops through the WSN, packet loss starts to take relevant values, although that only happens in clusters with assigned video MIP. Video MIP spends an amount of bandwidth between 10 and 50 times thee Audio MIP, so the probability of collision on the wireless network grows. When the number hops rises to three, significantly packet loss takes place even for audio MIPS with 64 Kbps and 192 Kbps. We can see that video delivery through three sensor nodes and 3500 Kbps bandwidth, equivalent to HDTV at 1080p, produces over 1% of packet loss. As a function of the codec used for video delivery, this percentage of packet loss can decrease drastically the quality of experience (QoE) of the end user.

The conclusion that we can extract from these results is that loss packet parameter can become a decisive QoS parameter that must be considered when the number of hops is equal to or bigger than three hops and the spent multimedia delivery bandwidth is high. MIP based cluster architecture can help by two ways: limiting the number of hops into a specific cluster and isolating heavy multimedia traffic into a separate cluster in order to improve QoS parameters of the other clusters.

6. Conclusion

Recently, the interest on WSNs has been increasing considerably, mainly because the nodes capacity to deliver huge amount of data efficiently in isolated geographic zones in harsh environments. The augmentation of the bandwidth in the new wireless technologies makes possible the use of multimedia sensors with new WSN usages. One of the main requisites in real time audio IP and video IP delivery is to meet the QoS requirements, but this is a difficult task in such types of networks. The way the WSN is organized and how sensor nodes communicate and create neighborhoods will be decisive to guarantee QoS.

In this work we propose and develop a new communication protocol that creates ad hoc clusters based on the multimedia flow features that are delivered inside the WSN. In order to achieve this goal, we have defined the MIP as a logical scheme that lets us manage the QoS requirements and the features of the sensor nodes building the cluster. The protocol allows the creation of clusters with a maximum diameter, which is adequate for each type of multimedia

Table 9: Statistical values and confidence interval for jitter as a function of the number of hops.

Hops	Parameters						
	N	μ (ms)	σ (ms)	Min (ms)	Max (ms)	Confidence interval (ms)	
1	1000	1.485	0.974	0	5	1.406	1.564
2	1000	3.118	1.336	1	12	3.009	3.227
3	1000	8.709	1.387	6	15	8.596	8.822

Table 10: Packet loss percentage.

Packet loss	AUDIO_64K	AUDIO_192K	VIDEO_1500K	VIDEO_3500K
1 Hop	0.00%	0.00%	0.00%	0.00%
2 Hops	0.00%	0.00%	0.04%	0.10%
3 Hops	0.19%	0.26%	0.51%	1.30%

flow and selects the most appropriate nodes, with enough resources, to be in the path of the multimedia delivery. We have detailed the protocol features, the designed messages, and the used variables. Moreover, we have explained the processes of the architecture, detailing how neighbour discovery, neighborhood creation, and multimedia delivery are taken place. Finally we have measured several cases in a test bench with real devices. We have proved that the protocol is able to achieve the adequate values of QoS parameters for different MIPs.

Our future research is focused on adding new MIP parameters such as sensor node mobility, energy consumption (like it has been added in [30]), and network stability [31]. Moreover, we are going to include the distribution capacity to the routing algorithm and add security mechanisms to guarantee the authenticity and integrity of the delivered multimedia data.

Conflict of Interests

The author(s) declare(s) that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been partially supported by the "Ministerio de Ciencia e Innovación," through the "Plan Nacional de I+D+i 2008–2011" in the "Subprograma de Proyectos de Investigación Fundamental," Project TEC2011-27516. This work has also been partially supported by the Instituto de Telecomunicações, Next Generation Networks and Applications Group (NetGNA), Portugal, by the Government of Russian Federation, Grant 074-U01, and by National Funding from the Fundação para a Ciência e a Tecnologia (FCT) through the PEst-OE/EEI/LA0008/2013 Project.

References

[1] D. Bri, M. Garcia, J. Lloret, and P. Dini, "Real deployments of wireless sensor networks," in *Proceedings of the 3rd International Conference on Sensor Technologies and Applications (SENSORCOMM '09)*, pp. 415–423, Athens, Greece, June 2009.

[2] M. Garcia, D. Bri, S. Sendra, and J. Lloret, "Practical deployments of wireless sensor networks: a survey," *International Journal on Advances in Networks and Services*, vol. 3, no. 1-2, pp. 170–185, 2010.

[3] L. Karim, A. Anpalagan, N. Nasser, and J. Almhana, "Sensor-based M2M agriculture monitoring systems for developing countries: state and challenges," *Network Protocols and Algorithms*, vol. 5, no. 3, pp. 68–86, 2013.

[4] J. Lloret, L. Shu, R. Lacuesta, and M. Chen, "User-oriented and service-oriented spontaneous ad hoc and sensor wireless networks," *Ad-Hoc and Sensor Wireless Networks*, vol. 14, no. 1-2, pp. 1–8, 2012.

[5] M. Edo, A. Canovas, M. Garcia, and J. Lloret, "Providing VoIP and IPTV Services in WLANs," in *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*, pp. 426–444, IGI Global, 2011.

[6] J. Mueller, T. Magedanz, and J. Fiedler, "NNodeTree: a scalable peer-to-peer live streaming overlay architecture for next-generation-networks," *Network Protocols and Algorithms*, vol. 1, no. 2, pp. 61–84, 2009.

[7] R. Diab, G. Chalhoub, and M. Misson, "Overview on multichannel communications in wireless sensor networks," *Network Protocols and Algorithms*, vol. 5, no. 3, pp. 112–135, 2013.

[8] L. Khoukhi and S. Cherkaoui, "Intelligent QoS management for multimedia services support in wireless mobile ad hoc networks," *Computer Networks*, vol. 54, no. 10, pp. 1692–1706, 2010.

[9] C. J. Barenco Abbas, L. J. García Villalba, and A. L. Sandoval Orozco, "A distributed QoS mechanism for ad hoc network," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 11, no. 1, pp. 25–30, 2012.

[10] Y. Chen, T. Farley, and N. Ye, "QoS requirements of network applications on the internet," *Information Knowledge Systems Management*, vol. 4, no. 1, pp. 55–76, 2004.

[11] T. Çevik and A. H. Zaim, "A multichannel cross-layer architecture for multimedia sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 457045, 11 pages, 2013.

[12] Z. Li, J. Bi, and S. Chen, "Traffic prediction-based fast rerouting algorithm for wireless multimedia sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 176293, 11 pages, 2013.

[13] J. Lloret, C. Palau, F. Boronat, and J. Tomas, "Improving networks using group-based topologies," *Computer Communications*, vol. 31, no. 14, pp. 3438–3450, 2008.

[14] J. Lloret, M. Garcia, J. Tomás, and F. Boronat, "GBP-WAHSN: a group-based protocol for large wireless ad hoc and sensor networks," *Journal of Computer Science and Technology*, vol. 23, no. 3, pp. 461–480, 2008.

[15] M. Lehsaini, H. Guyennet, and M. Feham, "An efficient cluster-based self-organisation algorithm for wireless sensor networks," *International Journal of Sensor Networks*, vol. 7, no. 1-2, pp. 85–94, 2010.

[16] J. Lloret, M. Garcia, D. Bri, and J. R. Diaz, "A cluster-based architecture to structure the topology of parallel wireless sensor networks," *Sensors*, vol. 9, no. 12, pp. 10513–10544, 2009.

[17] J. R. Diaz, J. Lloret, J. M. Jimenez, and S. Sendra, "MWAHCA: a multimedia wireless Ad Hoc cluster architecture," *The Scientific World Journal*, vol. 2014, Article ID 913046, 14 pages, 2014.

[18] D. Wei and H. Anthony Chan, "Clustering ad hoc networks: Schemes and classifications," in *Proceedings of the 3rd Annual IEEE Communications Society on Sensor and Ad hoc Communications and Networks (SECON '06)*, pp. 920–926, Reston, Va, USA, September 2006.

[19] J. Yu and P. Chong, "A survey of clustering schemes for mobile ad hoc networks," *IEEE Communication Surveys*, vol. 7, no. 1, pp. 32–48, 2005.

[20] R. Agarwal and M. Motwani, "Survey of clustering algorithms for MANET," *International Journal on Computer Science and Engineering*, vol. 1, no. 2, pp. 98–104, 2009.

[21] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826–2841, 2007.

[22] O. Boyinbode, H. Le, and M. Takizawa, "A survey on clustering algorithms for wireless sensor networks," *International Journal of Space-Based and Situated Computing*, vol. 1, no. 2-3, pp. 130–136, 2011.

[23] L. Ramachandran, M. Kapoor, A. Sarkar, and A. Aggarwal, "Clustering algorithms for wireless ad hoc networks," in *Proceedings of the 4th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM '00)*, pp. 54–63, August 2000.

[24] M. Chatterjee, S. K. Das, and D. Turgut, "WCA: a weighted clustering algorithm for mobile Ad Hoc networks," *Cluster Computing*, vol. 5, no. 2, pp. 193–204, 2002.

[25] M. Kavitha and K. Karthikeyan, "Comparison of data centric protocols for WSN and energy enhanced M-Spin (EEM-SPIN)," *International Journal of Engineering Research & Technology*, vol. 1, no. 9, 2012.

[26] Y.-M. Huang, M.-Y. Hsieh, and M.-S. Wang, "Reliable transmission of multimedia streaming using a connection prediction scheme in cluster-based ad hoc networks," *Computer Communications*, vol. 30, no. 2, pp. 440–452, 2007.

[27] S. Tang and W. Li, "QoS supporting and optimal energy allocation for a cluster based wireless sensor network," *Computer Communications*, vol. 29, no. 13-14, pp. 2569–2577, 2006.

[28] D. Rosário, R. Costa, H. Paraense et al., "A hierarchical multihop multimedia routing protocol for wireless multimedia sensor networks," *Network Protocols and Algorithms*, vol. 4, no. 4, pp. 44–64, 2012.

[29] J. R. Diaz, J. Lloret, J. M. Jiménez, and M. Hammoumi, "A new multimedia-oriented architecture and protocol for wireless Ad Hoc networks," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 16, no. 1, 2014.

[30] P. Spachos, P. Chatzimisios, and D. Hatzinakos, "Energy aware opportunistic routing in wireless sensor networks," in *Proceedings of the Global Communications Conference (GLOBECOM '12)*, pp. 405–409, December 2012.

[31] N. Meghanathan and P. Mumford, "Centralized and distributed algorithms for stability-based data gathering in mobile sensor networks," *Network Protocols and Algorithms*, vol. 5, no. 4, pp. 84–116, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 503417, 16 pages
http://dx.doi.org/10.1155/2014/503417

Research Article

WSN4QoL: A WSN-Oriented Healthcare System Architecture

S. Tennina,[1] M. Di Renzo,[2] E. Kartsakli,[3] F. Graziosi,[1] A. S. Lalos,[3] A. Antonopoulos,[4] P. V. Mekikis,[3] and L. Alonso[3]

[1] *WEST Aquila Srl, University of L'Aquila, 67100 L' Aquila, Italy*
[2] *Supelec, CNRS, 91192 Paris, France*
[3] *Department of Signal Theory and Communications (TSC), Technical University of Catalunya (UPC), 08034 Barcelona, Spain*
[4] *CTTC, Castelldefels, 08860 Barcelona, Spain*

Correspondence should be addressed to S. Tennina; tennina@westaquila.com

Received 11 December 2013; Accepted 17 February 2014; Published 6 May 2014

Academic Editor: Sana Ullah

People worldwide are getting older and this fact has pushed the need for designing new, more pervasive, and possibly cost effective healthcare systems. In this field, distributed and networked embedded systems, such as wireless sensor networks (WSNs), are the most appealing technology to achieve continuous monitoring of aged people for their own safety, without affecting their daily activities. This paper proposes recent advancements in this field by introducing WSN4QoL, a Marie Curie project which involves academic and industrial partners from three EU countries. The project aims to propose new WSN-based technologies to meet the specific requirements of pervasive healthcare applications. In particular, in this paper, the system architecture is presented to cope with the challenges imposed by the specific application scenario. This includes a network coding (NC) mechanism and a distributed localization solution that have been implemented on WSN testbeds to achieve efficiency in the communications and to enable indoor people tracking. Preliminary results in a real environment show good system performance that meet our expectations.

1. Introduction

Medical experts agree on the projections that forecast in the next couple of decades the population aged over 65 years to increase from 6.9% to 12.0% worldwide and, in particular, from 15.5% to 24.3% in Europe [1], with an average worldwide life-span expected to extend another 10 years by 2050 [2]. The growing number of older adults increases the demands on the public healthcare system and on medical and social services. Increased life expectancy reflects, in part, the success of public healthcare interventions [3], but public healthcare programs must now respond to the challenges created by this achievement, including the growing burden of chronic illnesses, injuries, and disabilities and increasing concerns about future care-giving and healthcare costs [4]. A study of Frost and Sullivan [4] has clearly indicated that, in almost all countries worldwide, healthcare spending per capita is rising faster than per capita income. If current trends hold, by 2050 healthcare spending is expected to double, claiming 20%–30% of gross domestic product (GDP) for some economies and 20% by 2015.

In this context, new technologies that can help seniors live at home longer provide a "win-win" effect, both improving quality of life and potentially saving enormous amounts of money. The forecast above clearly highlights the compelling need of delivering high-quality care to a rapidly growing population of elderly, while reducing the overall healthcare costs. Until recently, the cost of providing a continuous patient monitoring flow of patients' data, from patients' homes to care providers, was prohibitive, mainly because it requires continuous in-person patient monitoring through specially trained care givers available full time or dedicated communication and appropriate device infrastructures.

Nowadays, with the increasing availability of broadband technology at home, along with wireless networks and a wide range of consumer health electronics, an end-to-end

infrastructure has begun to emerge, enabling new "pervasive healthcare" applications, more often shortly termed as "e-Health" applications. The availability of these networks and the widespread use of mobile devices make two-way continuous interactions between patients and their care providers feasible, regardless of physical location.

On the other hand, the use of information and communication technologies to facilitate and improve healthcare and medical services involves the use of appropriate devices both on patients and embedded in the living environment. Researchers and educators predict that these gadgets will soon turn the home into a medical nurse, keeping record on everything from pill-taking routines to signs of imminent crises.

For seniors, the benefits of an idealized medical smart home are physical, psychological, and emotional. Aging in place means continuing with familiar routines while health data, detection of critical conditions, and remote control of certain medical treatments are wirelessly made available to doctors, caregivers, and concerned family [5]. For society, the bonuses include significantly reduced health care costs and happier elders. Consequently, it is natural to expect that consumers will embrace in-home technologies precisely because they can potentially save them money on the cost of unnecessary time-consuming doctor visits, thus not jeopardizing their daily life activities.

On this subject, this paper aims to present the WSN4QoL project [6], which involves the design of wireless sensor networks (WSNs) specifically suited to meet healthcare application requirements. WSN4QoL is a 3-year project started at the end of 2011 and still ongoing. With the intent of bringing together experts, from industry and academia, it proposes the use of advanced WSN technologies for pervasive healthcare applications. In particular, it aims at providing network coding (NC) for multihop/cooperative diversity in the data communication protocol as well as distributed localization algorithms to meet the specific requirements of WSNs-enabled healthcare applications, namely, energy-efficiency, low-latency, data reliability, context-awareness, and security. Extensive performance analysis of the proposed solutions is given through numerical simulations as well as proofs-of-concept in real-world experiments, through the implementation into real healthcare devices.

The rest of the paper is organized as follows. Section 2 overviews the full system architecture presenting the design challenges offered by the general reference scenario, while Section 3 puts particular emphasis on the key elements of the communication protocol stack and the middleware of services offered to the application designers. Sections 4 and 5 present the NC and the localization testbeds, respectively, that have been implemented to evaluate the solutions proposed in real(istic) environments. Section 6 deals with a summary of current research efforts with similar objectives to WSN4QoL. Finally, Section 7 concludes the paper with a view on some of the open issues to be addressed as ongoing and future work in the remaining WSN4QoL project period.

2. WSNs System Architecture Design for E-Health

WSNs are distributed networked embedded systems where each node combines sensing, computing, communication, and storage capabilities. They have emerged as a new networking environment that provides end-users with intelligence and a better understanding of the environment. Because of their wide variety of applications, it is envisioned that, in the near future, WSNs will become an integral part of our everyday lives [7]. WSNs are wireless ad hoc networks composed of inexpensive nodes with sensing capabilities and a limited number of data sink nodes. These nodes communicate among each other by forming multihop wireless networks and by maintaining connectivity in a centralized or a distributed manner. The network topology is, in general, dynamic, since the connectivity among the nodes may vary with nomadic and mobile nodes.

2.1. Challenges. Although fundamental research results on WSNs theory and practice have been achieved for many different applications, for example, traffic monitoring, plant monitoring in agriculture, and infrastructure monitoring, the application of this technology to e-Health poses some unique application-specific challenges and constraints. In particular, the efficient design of a WSNs-enabled pervasive healthcare system is characterized by the following intrinsic differences with respect to "general-purpose" WSNs design, which require special attention [8]. (i) The devices have limited available energy resources, as they have a very small form factor. (ii) A low transmit power per node is needed to minimize interference and to cope with health concerns. (iii) The devices are located on the human body, which can be in motion. WSNs for e-Health should therefore be robust against frequent changes in the network topology and channel variability. (iv) Data mostly consist of medical information; hence, high reliability and low delay/latency are required. (v) Stringent security mechanisms are required to ensure the private and confidential character of data. (vi) Context-awareness through cooperative localization in outdoors and indoors is crucial to enable a prompt reaction in case of emergency. (vii) The devices are, in general, very heterogeneous. They may have very different demands or may require different resources of the network in terms of data rates, power consumption, and reliability.

Along these lines, the main research objective of WSN4QoL is to provide fundamental research advances, proof-of-concepts, and real-life implementations on the main enabling technologies for WSNs-aided e-Health applications. More specifically, disruptive techniques such as cooperative wireless communications protocols and distributed algorithms are investigated. The proposed solutions are designed, optimized, and implemented in real-devices by taking into account the specific requirements of e-Health: energy efficiency, low-latency delivery of data, data reliability, and security. In more detail, the research objectives of WSN4QoL include the following: (i) to design a protocol stack architecture, which can accommodate a variety of protocols,

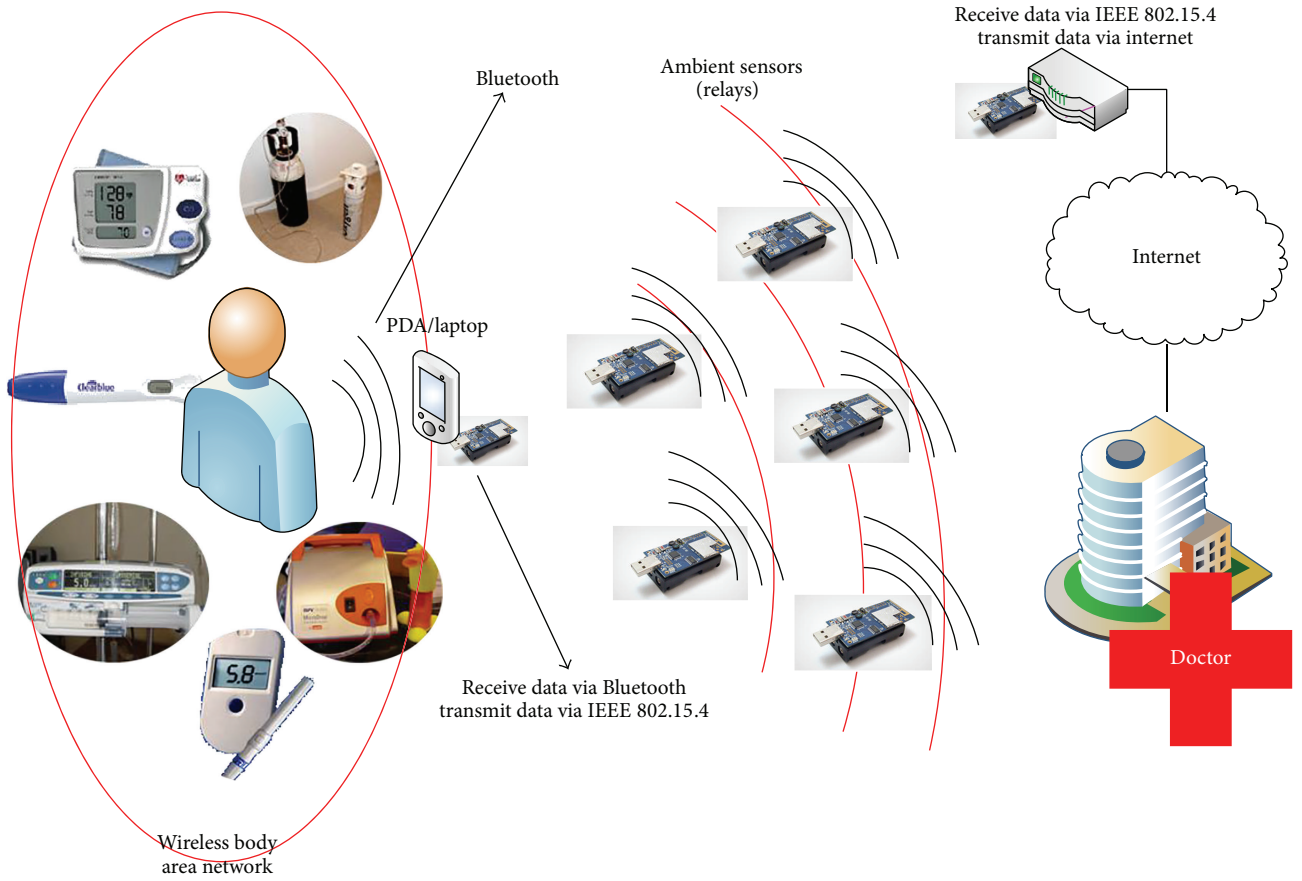


Figure 1: Reference healthcare 3-tier system architecture.

algorithms, and sensor devices for pervasive healthcare applications, (ii) to develop energy-efficient and performance-guarantee cooperative protocols and NC schemes for WSNs-enabled pervasive healthcare applications, (iii) to propose advanced distributed localization protocols and algorithms specifically suited for the scenarios (e.g., indoors) envisaged by WSNs-enabled pervasive healthcare applications, (iv) to conceive effective, efficient, and resilient security solutions for the proposed algorithms and protocols, (v) to implement and assess the performance of the protocol stack in a WSN testbed, and (vi) to integrate the proposed solutions in real devices and validate them in real working environments.

2.2. Reference Scenario. Similar to other works in literature (e.g., [9, 10]), the reference system architecture proposed in this project is as depicted in Figure 1. It is a three-tier system architecture, where at the lowest tier (*Tier-1*), a Bluetooth-enabled WBAN connects sensors to a local collector (i.e., a hub), which can be a portable embedded PC or a PDA. The hub needs to communicate with WBAN devices through a Bluetooth radio module and then send measurements reports towards a residential gateway, through a ZigBee/IEEE 802.15.4 based multihop WSN (*Tier-2*). The gateway is able to perform local computation and forward data to the public IP-based network (*Tier-3*) towards the professional caregivers for real-time analysis.

In recent work [11–13] we proposed alternatives to the Bluetooth for the communications among the devices forming the WBAN at the Tier-1. Nevertheless, in the WSN4QoL project, our focus is on the efficient data transmission over the WSN network at the Tier-2, as well as supporting real-time people localization in a fully distributed way.

3. System Protocol Stack

Figure 2 shows the intended communication protocol stack for the WSN of the reference scenario in Figure 1. Moving from the bottom, the protocol stack is composed of the following entities.

(a) IEEE 802.15.4 MAC Layer. This layer is responsible for the access to the wireless medium for transmission and reception of the frames for both mobile patients and fixed relay nodes.

Among the options offered by the IEEE 802.15.4 standard [14], we have chosen to refer to the *non-beacon-enabled mode*, that is, the fully asynchronous mode. This choice is motivated by the fact that nodes can have variable duty cycle, especially the mobile ones, that is, those carried by the patients. In the classical scenario where patient's data need to be collected at a central station, the asynchronous mode offered by this standard allows for flexibility in accessing the medium only when patients' data are available and a transmission needs

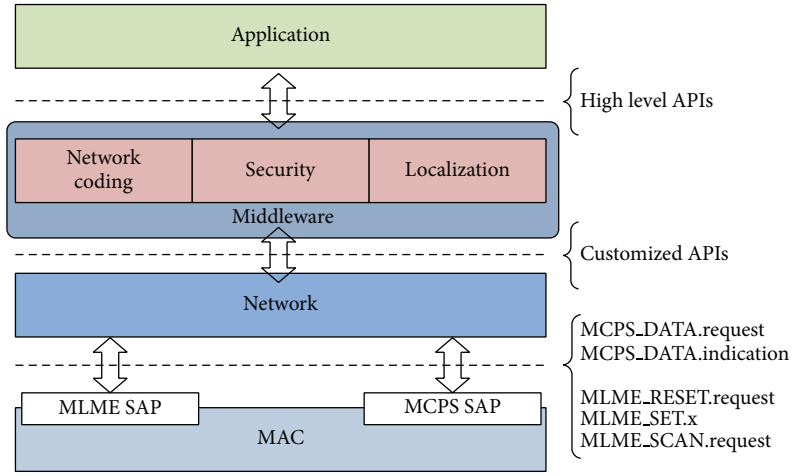


Figure 2: Communication protocol stack on top of the IEEE 802.15.4 standard. Basic interfaces are also shown.

to occur. For the majority of time, the radio interface of these nodes can be kept off or in a low-consumption state to save the energy of the batteries. Moreover, the synchronous mode would have required association (and disassociation) mechanisms to allow the nodes to join the network prior to communication, resulting in a severe limitation on the nodes' mobility. Furthermore, to exploit the power of NC mechanisms for energy savings and network throughput gains (Section 4), messages sent by the mobile nodes need to be transmitted in broadcast, thus without a prior association mechanism.

Finally, along this line, the basic commands and events offered by the MAC to the upper layer are for packet transmission and reception and to set some specific parameters, such as the frequency channel and the transmission power, as well as the primitive to scan the IEEE 802.15.4 channels for energy and activity detection.

(b) Network Layer. Since the asynchronous mode of the MAC layer is chosen, this layer is responsible for keeping the synchronization among the fixed relay nodes. This is achieved by sending *synch* packets. Unlike IEEE 802.15.4 beacon frames, synch messages are not requested to be periodic: their transmission can be scheduled with an adaptive duty cycle, based on the environmental conditions (e.g., the presence of patients in the area or not), although for keeping their scheduling a mechanism inspired by the time division cluster scheduling [15] can be implemented.

Synch packets are fundamental to allow for minimizing the collisions among the messages sent by the mobile nodes, by defining a *superframe* structure constituted of time slots, where each mobile node is allowed to transmit, based on some policy rule. Finally, they are requested to implement both NC schemes and distributed localization algorithms, as will be detailed in the next sections.

Besides the synchronization, the network layer is also in charge of assigning the network addresses to the nodes. Usually, the radio interface of a node has an address which is worldwide unique as a serial number assigned by the manufacturer. In the case of IEEE 802.15.4 radios, these addresses are 64 bits long and can be used for the communication with any other node (extended address). Another option is for a node to get a network address, which is only 16 bits long, assigned according to some policy and used to communicate with the other nodes of the same network(short address).

Shortly, the extended address mechanism is used by the mobile nodes. The fixed relays are usually placed in strategic positions to ensure the best coverage of the environment and form a network with a static topology. Consequently, these nodes can be assigned with the addresses defined by, for example, the ZigBee Distributed Address Assignment Mechanism. This addressing mechanism assumes that nodes are organized into a tree and divides the address space into blocks, assigning each block to each node of the tree. The advantage of using short addresses in this way for the relay nodes is the fact that they do not need to maintain routing tables to forward incoming data: simply looking at the address, they are able to recognize if the packet has to be sent upwards or downwards along the tree. Finally, also packets to the mobile nodes are assumed to be sent in broadcast: this is still reasonable since it is assumed that the network traffic from the residential gateway to any patient (related to any actuation, such as, for example, automatic regulation of an insulin pump) is less frequent than the reverse direction (patient's monitored data collection).

The commands and events offered by this layer to the upper modules are customized based on the reference model we assume (ZigBee) with the addition of those elements needed to implement the services at the middleware layer.

(c) Middleware Services. This layer represents an interface between the underlying protocol stack and the application layer and is the core of the novelties introduced within the frame of WSN4QoL project.

The middleware encompasses three major blocks: (i) NC; (ii) distributed localization; (iii) security. These blocks exploit the services offered by the underlying protocol stack entities

to provide a high-level application programming interface (API) to the application developers.

The NC entity is in charge of providing efficiency in wireless communications. By means of appropriate combinations of two or more packets into a single one and NC-aware routing mechanisms at the lower layer, a relay node is able to reduce the amount of traffic over the network, without losing data. Section 4 will present the basic building blocks developed within WSN4QoL to demonstrate the efficiency of a binary XOR-based network coding scheme in a scenario with two sources, one relay and one destination (i.e., a *multiple access relay channel* scenario), as compared to the case where the relay node simply forwards the received packets. In general, the proposed protocol stack allows for adopting multicast (and geographic) routing mechanisms at the network layer [16, Chapter 6] which are well suited for supporting NC schemes more complex than the binary XOR-based ones, including, for example, the random linear network coding (RLNC) [17].

The distributed localization block deals with the online estimation of the geographical position of a mobile node in the environment. Associating a spatial reference with every communication between the patients and the remote care givers is of paramount importance, whether in home or hospital environments, especially in case of alarms conditions. As better detailed in Section 5, the relay nodes emit their *synch* packets and doing so they play the role of *anchor* or *reference* nodes; that is, it is supposed they know their own position and they include this information in such packets. A mobile node is then able to estimate its own position by relying on the information gathered from the surrounding nodes.

The security block monitors the acknowledgement packets exchanged at the network layer among the nodes to identify potential threats or nodes malfunctioning and instruct the MAC layer to encrypt frames based on the security features offered by the IEEE 802.15.4 standard. Details about this block are out-of-scope of this paper, where the focus is switched to the other two elements, described in the next sections.

(d) Application Layer. This last layer mainly focuses on gathering measurements from the sensors and data compression. Although the WSN4QoL project also proposes low cost compress sensing techniques which exploit some key characteristics of the biometric data transmitted in order to provide energy-efficient telemonitoring solutions, this layer is out-of-scope of the present paper. Interested readers can find further details in our project website [6].

Next sections will present the implementation and experimental results of a binary XOR-based NC scheme and the distributed localization algorithm. Table 1 summarizes the main features of the WSN platforms used for the tests.

4. Network Coding

To achieve efficient measurements reporting through the ambient relay network, the most viable solution is the application of NC techniques [18, 19]. In the preliminary implementation done in the frame of the WSN4QoL project,

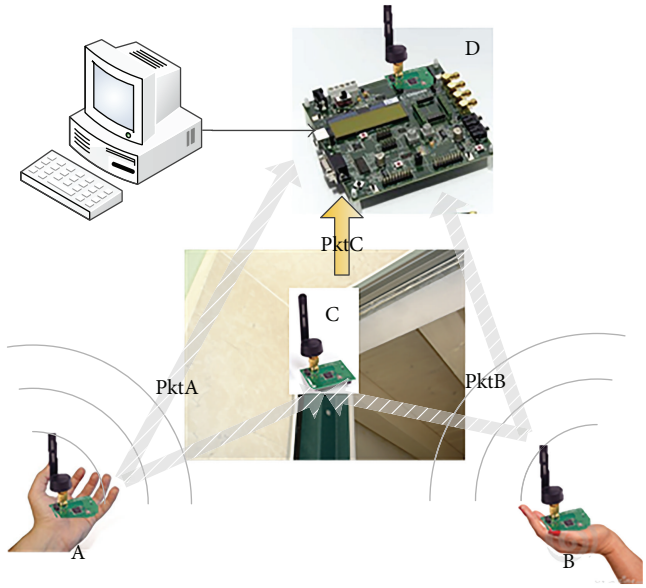


Figure 3: Scenario for efficient communications. Nodes A and B are mobile local hubs, node C is a fixed relay, and node D is the destination.

the scenario illustrated in Figure 3 has been implemented in the available testbed. In this scenario, two nodes A and B are mobile nodes carried by two patients; a relay node C has a fixed position and D is the destination of the measurements reports that are sent by the two source nodes. In the considered scenario, to further claim the gains introduced by the NC techniques, it is supposed that the destination node does not send back any feedback either to the relay or to the sources.

4.1. Implementation. Figure 4 presents the timings characterizing the scenario of Figure 3. In particular, Figure 4(a) presents the case of the baseline scenario and Figure 4(b) the case with NC.

The relay C is responsible for defining the *network scheduling* among the source nodes A and B by periodically broadcasting the synch packets. The source nodes are programmed to send a message in their appropriate time slots; that is, node A sends its data after Tslot and node B sends its data after 2*Tslot. In the baseline scenario, node C then forwards the received data from A to the destination after Tslot and data from B after another Tslot; then it broadcasts a new synch packet and the process is iterated. In the NC scenario, instead of forwarding the two messages, the node C transmits to the destination a single message which is the combination of the two messages from A and B based on a binary XOR operation. In both cases, upon the reception of every synch, the destination checks what it has received in the *superframe* just concluded and updates the network statistics.

The main advantage of the NC against the baseline scenario is that the relay node forwards a single packet instead of two, resulting in larger energy savings and the higher throughput since the synch is sent every 4*Tslot instead of 5*Tslot.

Table 1: WSN platforms testbeds features.

Item	Specification		Description
	TelosB	CC2430	
	Processor		
Model	Texas Instruments MSP430F1611	Intel 8051	
	16 bit	32 bit	
Memory	48 KB	128 KB	Program flash
	10 KB	8 KB	Data RAM
	1 MB	—	External Flash
ADC	12 bit 8 channels		
Interface	UART, SPI, I2C	Two programmable USARTs	
	USB	21 GPIO	
	Radio		
RF Chip	Texas Instruments CC2420		IEEE 802.15.4 2.4 GHz Wireless Module
Frequency Band	2.4 GHz–2.485 GHz		IEEE 802.15.4 compliant, ch11–ch26
Sensitivity	−95 dBm typical		
Transfer Rate	250 Kbps		
RF Power	−25.2 dBm–0.6 dBm		Software Configurable
Range	120 m (outdoor) 20–30 m (indoor)		
Current Draw	RX: 18.8 mA TX: 17.4 mA Sleep mode: 1 uA		
Antenna	PCB Antenna	Dipole Antenna	

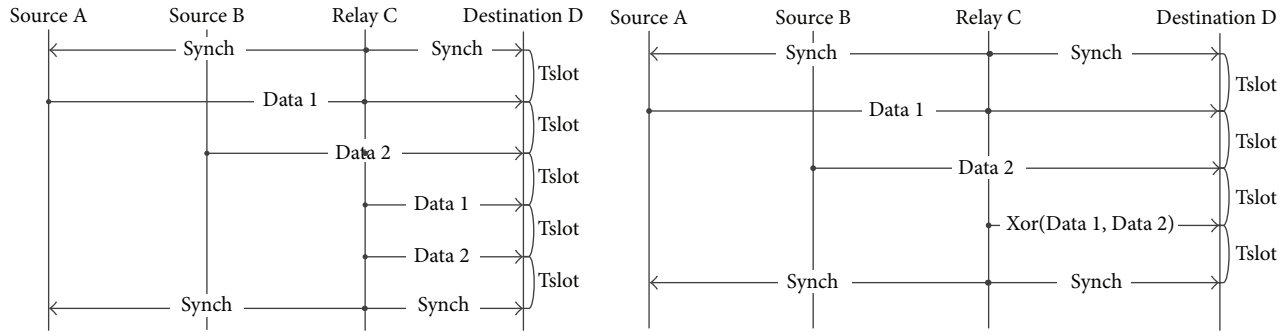
Both scenarios have been coded and tested over a WSN testbed composed by TelosB nodes [20] (Table 1), programmed with the TinyOS operating system [21], and implemented as protocol layer(s) on top of the Official TinyOS 15.4 MAC [22, 23]. The nodes of the baseline scenario have been configured to run on the channel 25 of the IEEE 802.15.4 standard, while the nodes of the NC scenario run on the channel 26: this is done in order to have the two networks working at the same time and test them in the same conditions.

To monitor in real time the network behaviour and the performance, a graphical user interface (GUI) has also been implemented and is shown in Figure 5. It shows (i) the nodes' transmissions, (ii) the values of the sensors readings, and (iii) the residual of the batteries which are data encoded in the transmitted packets, as well as the performance metrics of the network including (iv) throughput or goodput (i.e., the amount of sensors measurements successfully delivered to the destination in the unit of time), (v) packet loss, and (vi) energy consumption.

4.2. Results. Compared to the classical relay scenario where node C forwards the received packets in two distinct slots, the NC scheme allows for achieving better performance in terms of joint packet loss ratio (PLR) and data goodput. In particular, the scenario of Figure 3 implemented on the mobile WSN testbed as described in the previous section was used in an indoor environment, such as a residential apartment, to run several tests by varying the transmission power levels between −25.2 dBm and 0.6 dBm. While source 2 nodes of both testbeds were kept in a fixed position, source 1 nodes of both networks were carried by a person who was walking at approximately constant speed over a preplanned closed path crossing the rooms of the apartment, resulting in a time of a lap of around 5 minutes and repeating the path at least 10 times for each experiment.

Although the PLR shows similar performance between the scenario with NC and the one with classical forwarding, the tests demonstrated that NC can achieve gains ranging from 32% to 68% in terms of instantaneous goodput. In particular, Figure 6 reports the goodput averaged over the whole experiments (i.e., all the laps for each transmission power) and for different values of the Tslot ranging from 200 ms to 800 ms (Figure 4). As it is evident, the NC shows always gains with respect to the baseline (relay-only) scenario for all the values of Tslot and all the transmission power levels. Moreover, analyzing the behavior with respect to the transmission power, both testbeds show similar performance and this further confirms that the two testbeds result in similar performance in terms of PLR.

Although in this preliminary testbed the simple XOR-based NC scheme has been implemented over a single relay scenario, future ongoing activities are focused on the enhancement of this mechanism in the cases where the ambient network is composed by several relay nodes, including multihop communications [24].



(a) Baseline scenario: relay forwards two messages

(b) Network coding: relay broadcasts a single message

FIGURE 4: Timings of the multiaccess relay channel scenario as depicted in Figure 3.

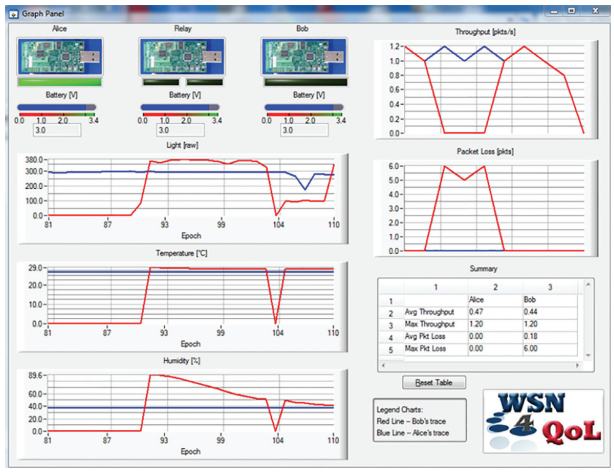


FIGURE 5: Graphical user interface to monitor network performance for network coding testbed.

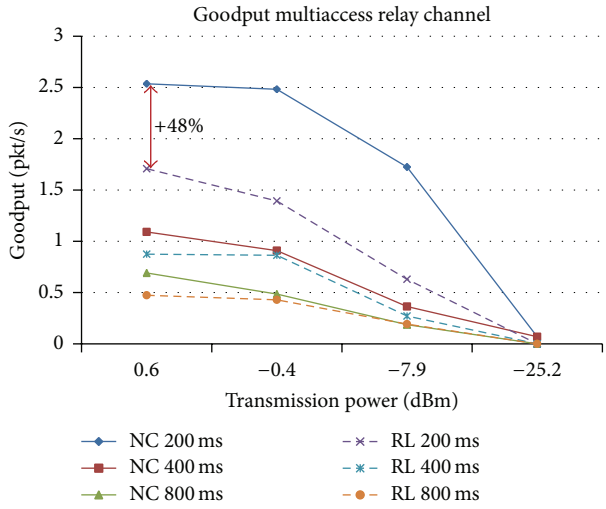


FIGURE 6: Goodput comparison in multiaccess relay channel scenarios: with network coding and relay only.

5. Distributed Localization

The aim of this section is to present the ongoing activity for indoor people localization in the WSN4QoL project. The algorithm we refer to is detailed in [25]. It is an *anchor-based* algorithm, which means it runs in a scenario where several fixed anchor nodes, that is, nodes knowing *a priori* their positions based on a common reference system of coordinates, are deployed in the environment and periodically broadcast their positions. A second set of nodes is mobile and called blind since they need to estimate their own positions according to the same reference system of coordinates, by relying on the data they are able to gather from the anchors and the environment. The algorithm is also *range-based*, since the blinds estimate their positions by first computing the distance with respect to the anchors available.

5.1. Implementation. Figure 7 shows a typical home environment where a set of anchor nodes are fixed and have been deployed in the rooms of the house. In such an environment a measurement campaign of the received signal strength (RSS) from the anchors in several points has been conducted, with the intention of building an RSS-to-distance relation curve, that is, an RSS-based ranging model used to estimate the distance between any pair of nodes from an RSS measurement.

Typically, the calibration activity to compute the parameters of the model is performed offline and in a static context; then the system needs to run in an environment which is more dynamic, for example, with people moving around or changes in the furniture and so on. As a consequence, RSS propagation parameters are strongly environment dependent and usually show big fluctuations, which suggest that using any fixed and outdated estimate for the channel parameters certainly yields less accurate estimates of distances and thus of final positions. To cope with this problem, an *anchor-aided* dynamic and adaptive estimation of the signal propagation parameters has been previously proposed and can be easily implemented [26] as shown in Figure 8.

In particular, the relays-anchors put in the synch packets the RSS received from other anchors in their communication

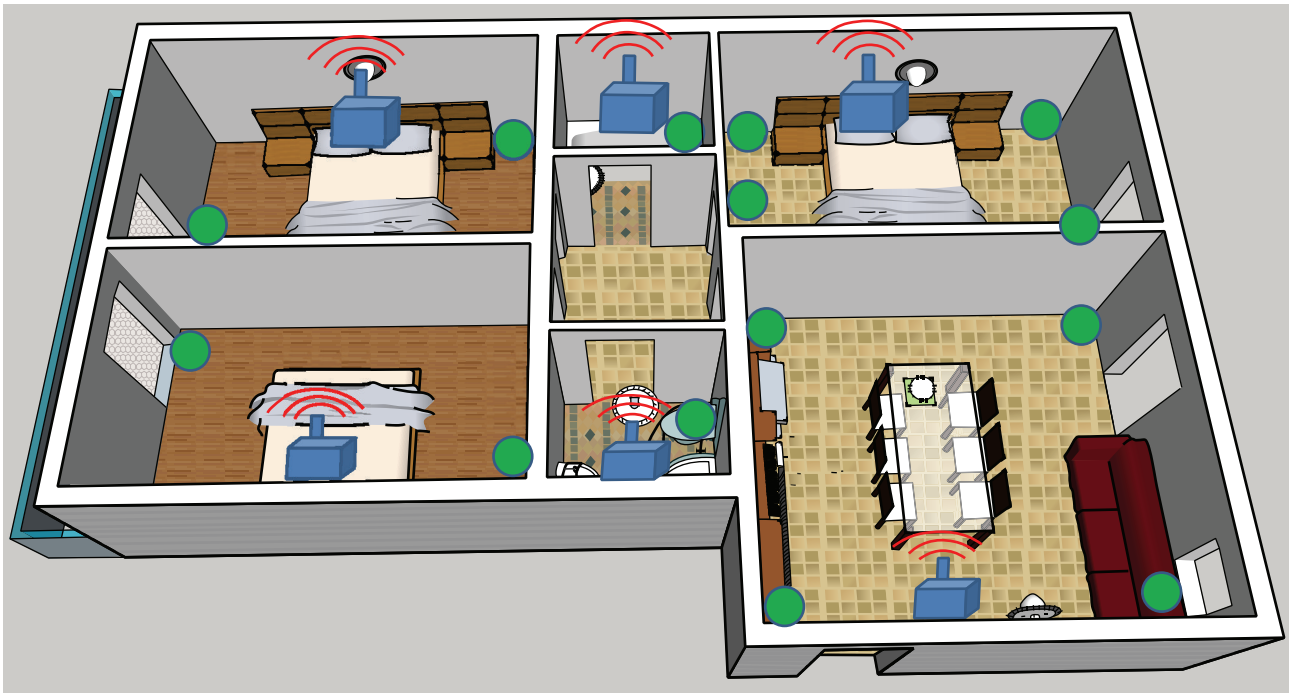


Figure 7: Indoor environment equipped with anchor nodes (blue squares) and RSS measurement points (green circles).

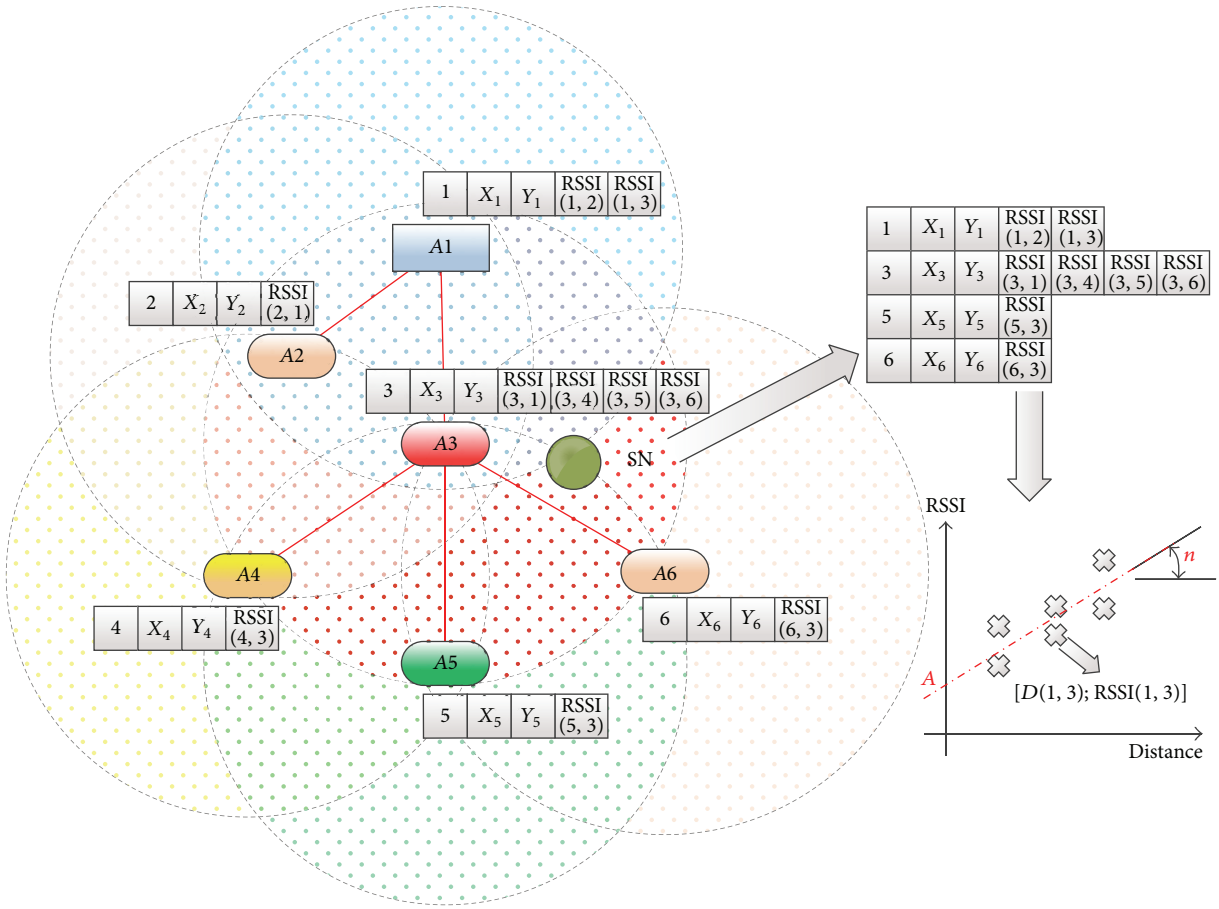


Figure 8: *Anchor-aided* dynamic and adaptive estimation of the ranging model.

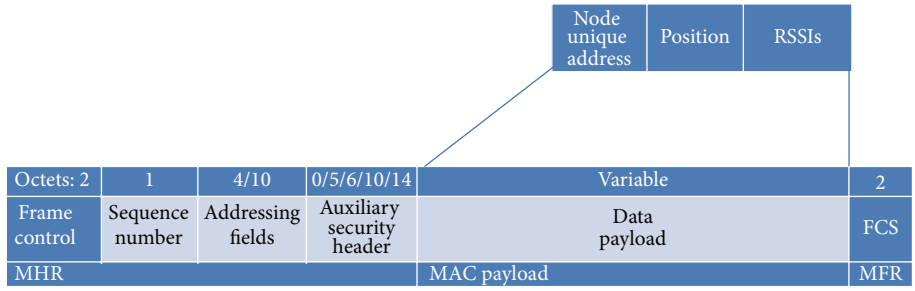


Figure 9: IEEE 802.15.4 data frame used by anchors with localization-oriented payload.

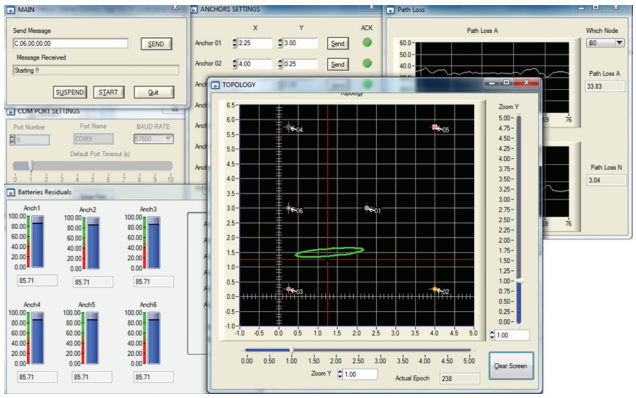


Figure 10: Graphical user interface to monitor network performance for Localization testbed.

range. Figure 9 shows how the IEEE 802.15.4 data frame is used for the anchors to transmit their positions, including also the RSS data needed for implementing the dynamic and adaptive ranging model estimation.

Every mobile node in the area receives these packets and is then able to correlate the distance among the available and known anchor nodes with their respective RSS so that the ranging model can be reconstructed as formulated in [26].

The implementation has been done on TelosB nodes [20] running the TinyOS operating system [21] for the anchor nodes and an IEEE 802.15.4-compatible platform for the mobile node, such as the Texas Instruments CC2430 [27] which has the same radio interface of the TelosB nodes (Table 1).

Similar to what has been done for the NC testbed, to monitor in real time the network behaviour and the performance, a GUI has been implemented and is shown in Figure 10. It allows for configuring at deployment time the position of the anchor nodes and then shows (i) the nodes' activities, (ii) the residual of the batteries which are data encoded in the transmitted packets for every node, and (iii) the statistics of the localization estimations for the mobile node such as (iv) instantaneous position estimation, (v) covariance ellipse (with a 70% confidence interval) of the last 5 estimations, and (vi) the ranging parameters estimated by the mobile node.

5.2. Results. Figure 11 presents preliminary results of the localization of a blind node placed in several points in the environment of Figure 7. The blind node has been left in each position for 5 minutes, resulting in at least 100 localization estimations.

It is evident that better topological conditions (i.e., where surrounded by the anchors) lead to better localization accuracies and stability (i.e., little average errors and low variability among estimations). Overall, the average localization error over the area of 60 m^2 is below 2.5 m, which results in a room-level accuracy which matches the requirements for e-Health applications. Nevertheless, there exist few critical situations on the borders. These issues can be solved by improving the coverage of the anchor nodes. Along this line, radio propagation simulation software (e.g., [28]) will provide the optimal anchors' number and positions.

6. End-to-End Solutions and Testbeds

After having presented the solutions proposed within the WSN4QoL project, this section overviews the state of the art in the field of WSN-based systems for pervasive healthcare applications, pointing out the main innovation aspects that the WSN4QoL project proposes.

6.1. Motivation. Before proceeding to give an overview of the solution and testbed in literature, an interesting vision of e-Health remote monitoring systems as given in [29, 30] is worth mentioning. In these works, authors classify the telecare applications as an instance of the broader *cyber physical world* (CPW), where a tight integration of sensing, computation, and communication elements concur to the definition of the system. In this line, several exciting research challenges and opportunities arise and might stimulate new research activities in the emerging areas of CPW convergence.

In general, nowadays, these technologies still require the development of reliable, scalable, and evolvable systems in various application domains. They should hide unnecessary complexities inherent to CPW, such as heterogeneity and distribution, and support rapid implementation of application and runtime reconfiguration and resource management to meet functional and nonfunctional requirements. One of the

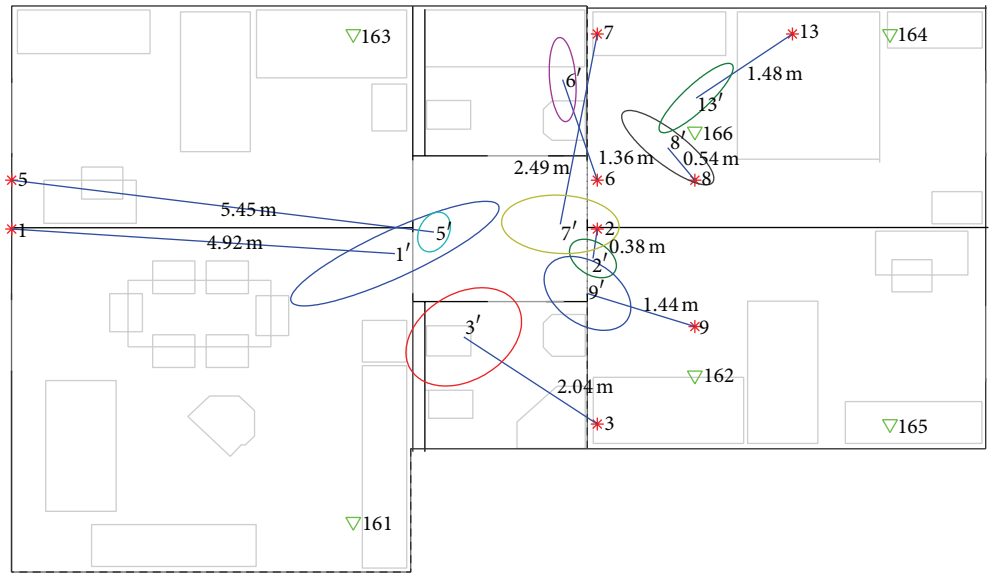


Figure 11: Localization results in the scenario of Figure 7. Average localization errors and covariance ellipses are also shown.

key elements focuses on *the study of methods and techniques that can be used to investigate the structure and evolution of the dynamics of human behavior under the lens of pervasive computing*. In this perspective, the sociotechnical nature of the CPW convergence calls for novel and interdisciplinary research approaches mixing ICT (information and communication technologies) expertise with lessons learned from applied psychology, sociology, medicine, complexity science, and so forth.

Along this line, wearable computing (WC) for remote assistance is one of the best examples where this multidisciplinary approach takes the most promising advantages. The maturity of WC can be demonstrated by several projects showing its main application domains. Monitoring the wearer's vital signs promises improved treatment and reduction of medical costs. Many projects are aiming at a preventive lifestyle and early diagnosis, by focusing on the integration of healthcare more seamlessly into everyday life.

Nevertheless, [29] stresses the fact that several research challenges have still to be met in this field. In particular, the large-scale production of smart clothes requires scaled-up manufacturing processes to exploit economy of scale effects. In addition, web-based data gathering methods (e.g., [31, 32]) are supportive in fusing heterogeneous sensor modalities and in the automatic annotation of data streams. Moreover, major challenges arise for data privacy and security in designing algorithms and protocols to protect the sensed data against unwanted collection and distribution of personal information.

The work in [33] summarizes the most recent developments in the field of wearable sensors and systems, relevant to the field of health and wellness, safety, home rehabilitation, assessment of treatment efficacy, and early detection of disorders. The integration of wearable and ambient sensors is discussed in the context of achieving home monitoring of older adults and subjects with chronic conditions. Particular emphasis is given to the analysis of the key enabling technologies to broaden the wearable systems for patients' remote monitoring. Among these, we underline the new advances in the sensors miniaturization's technologies and low-power microcontrollers and radios, which can be integrated into system-on-chip implementations and enable e-textile based systems (such as [34]) or unobtrusive wearable devices (such as [35]). Mobile phone technology played the major role in the pervasiveness of remote monitoring systems, since smart phones are broadly available and easily act as information gateways to central stations through mobile telecommunication standards such as 4G. Moreover, mobile devices can also function as information processing units. Reference [33] concludes that the past and ongoing research toward achieving remote monitoring of older adults and subjects undergoing clinical interventions will soon face the need for establishing business models to make them effective in the market to cover the costs. However, we believe that also a new area of research should involve the definition of new communication standards for the interoperability among the plethora of authorities, as stressed by similar other surveys, such as [36].

On the other hand, [37] presents a range of wireless communication technologies and standards (i.e., IEEE 802.15.1 Bluetooth, ZigBee/IEEE 802.15.4, Ultra-wide-band (UWB), medical implant communication services (MICS), and IEEE 802.15.6 Task Group). It lists their current limitations, by dividing the platforms into implantable and on-body devices. In particular, the integration of wireless technologies into medical devices, such as insulin pumps and on-body cardiac monitoring devices, has many benefits, but also poses many challenges, including (i) enabling the secure transmission of the collected private data, (ii) prevention of electromagnetic interference between different wireless devices and compatibility with the remaining circuitry, and (iii) compatibility with, and safety of, the biological tissues.

In this perspective, although security and privacy are well investigated in the literature and well taken into account in the projects, we believe that the latter two issues are still too often widely neglected. Of particular interest is the constraint posed by the last one. Besides the power and small size requirements, implantable devices also need to be compatible with biological tissues in order to prevent possible infection and rejection of the device by the biological tissue. Compared to implantable devices, on-body devices are less prone to the biocompatibility constraint. However, it is worth noting that also in this case long-term skin contact with such devices can cause different forms of skin irritations. Thus, on-body devices should also either be developed with biocompatible materials, or be truly noninvasive, where no skin contact is required for the acquisition of the desired data.

6.2. Testbeds. The authors in [38, 39] provide two examples of ECG telemonitoring implementation testbeds. Reference [38] focuses on the extension of the Standard Communication Protocol for Computer-assisted Electrocardiography (SCP-ECG), which provides standardized communication among different ECG devices and information systems, to be included in health monitoring systems. The paper describes the implementation of the new protocol as a software component in a five-month pilot period of health telemonitoring system (HTS) including 27 patients. The testbed they used to show the feasibility of their proposed enhancements includes all the elements of a telecare system. In particular, a wearable data acquisition system, consisting of several sensors (e.g., ECG, NiBP, SPO_2, Pulse Rate, Temperature, and PLE), is equipped with a Bluetooth radio, a GPS receiver, and a personal digital assistant (PDA) with mobile ADSL capabilities. The PDA automatically organizes the data gathered from the sensors and other information manually inserted by the user into the data structure defined by the protocol. A remote health monitoring system (RHMS) is implemented on a PC on the expert's site and is able to store, present, and process the acquired data from the PDA. The PDA communicates with the RHMS using fixed and mobile ADSL.

Although the suitability of the protocol has been clearly shown, security aspects are merely taken into account by the use of data encoded transmissions, while user identification of the involved people (i.e., individuals, technicians, physicians, etc.) is performed through a simple log-in screen requiring user ID and password.

On the contrary, [39] explicitly focuses on the implementation of security techniques in similar ECG-based telecare applications even if the scope of the paper is limited to the sensing device. Namely, a secure cross-layer-based miniaturized BSN platform has been developed. It consists of a processing unit and a radio transmission unit with a sensor board and a local battery power supply or energy scavenge supply. The design of such platform puts particular emphasis on resource-awareness; that is, it adopts a joint unequal resource allocation (i.e., transmission power and data rate) and real-time selective encryption, according to the channel status.

In [40] a body posture model and an unsupervised learning and clustering algorithm have been proposed to reconstruct different stationary postures. An extensive validation has been performed through a BSN composed by Freescale nodes [41] equipped with 3-axis accelerometers. These nodes are firmly attached through bands to four limbs to measure the posture of arms and legs, with two accelerometers on each limb, and report through a wireless single-hop ZigBee radio the measurements to a central station. Experimental results demonstrate that the proposed system can achieve very high classification accuracy and is able to recognize complicated stationary postures.

The authors in [42] use a pair of Shimmer motes [43]. Shimmer is a wireless sensor platform programmed in TinyOS [21], characterized by a small form factor, that can record and transmit physiological and kinetic data in real time using the most well-known communication technologies, such as Bluetooth or IEEE 802.15.4. The chosen device incorporates a triaxial accelerometer, a microcontroller, and an IEEE 802.15.4 radio transceiver. One mote is used as a wearable device while another is attached directly to a PC acting as a base station. Since the proposed method to extract features from acceleration measurements is not computationally intensive, the filtering technique has been implemented directly on the wearable device in order to communicate with the base station only when alarms occur and then save batteries.

In a slightly different scenario, that is, military missions and monitoring of soldiers, the contribution of [44] is the concept and implementation of a closed-loop, end-to-end, real-time on-body prediction system for reducing health risks due to uncompensable heat stress (UHS). This involves gathering physiological data (multipoint skin temperature) and postural information (multipoint body acceleration) for the purpose of autonomous real-time modelling and prediction. One of the central concepts driving this system development is that data processing must be performed by system devices mounted on the body to achieve a better real-time closed-loop control. Autonomous operation is essential because a long-range radio link to a central location might not necessarily be available. A relatively powerful hardware platform is thus required to support real-time on-body data processing, which also enables two control loops. An inner loop implements local actuation, namely, notifying alarms to the user or automatically taking some actions such as cooling the body. An outer loop involves the communication with a remote station for, for example, mission plan change or the return to the base to install new cooling systems. Similar approaches can be easily implemented in more traditional civil scenarios for remote patient monitoring.

Another simple yet efficient Internet-based telecare remote monitoring system is presented in [45], where the focus is on a remote-controlled home mechanical ventilation (HMV) system, which is progressively being used to treat patients with severe chronic respiratory failure. Contrary to the most conventional settings, the system designed avoids any high order information technology architecture. It is based on a simple and low cost data transfer server (DTS) that grants the Internet connection to most commercially

available ventilators through the GPRS network. The device captures ventilation signals (e.g., pressure, flows, volume, leaks, and oxygen saturation) and controls the ventilator settings. The DTS is built upon an embedded system board (a Linux-enabled FOX board [46]) equipped with a GPRS modem and a commercial USB flash memory. It operates as a web server with its own address and password. With such an approach, an independent point-to-point (from patient home to HMV provider) communication is established. Therefore, the HMV provider (being a hospital service or a private practice physician) can receive real-time or previously recorded ventilation data in uplink and modify the settings in downlink by simply connecting, via Internet, to the individual web address of the DTS at the patient's home, ensuring this way the closed-loop control.

A remote and mobile patient monitoring service architecture using heterogeneous wireless access in which each patient is equipped with a remote monitoring device with a heterogeneous wireless transceiver is presented in [8]. While the system architecture is not a novelty, since authors propose that a mobile patient can use different types of wireless technologies (e.g., WiMAX-based WMAN and WiFi-based WLAN technologies) to transfer monitored biosignal data to the healthcare center, the most innovative aspect in this contribution is the formulation as a constrained Markov decision process (CMDP) of the problem on the e-Health service provider side, who has to pay to the wireless network service provider a certain number of connections to be *reserved* for the patients. Using stochastic programming techniques, the optimal number of reserved connections can be determined to minimize the cost of the e-Health service provider under randomness of connection demand due to the mobility of the patients. Also, at the patient-attached device, the transmission scheduling of biosignal data with different priority is optimized to minimize the connection cost and satisfy the delay requirements.

In the frame of coupling wearable BANs with classical WSNs for environmental monitoring, [9] presents a study of a healthcare architecture for monitoring elderly or chronic people in their residence. Figure 12 sketches the reference network architecture, where wearable sensor system (composed by a single belt of sensors) communicates with powerful mobile computing devices through Bluetooth and the wireless sensor network through ZigBee. The higher part of the architecture includes communication technologies such as cellular-based networks and WiFi (or even WiMax). Although similar to WSN4QoL's reference scenario sketched in Figure 1, this architecture differs from that since patients sensors and ambient living WSN nodes form two distinct subnetworks. Nevertheless, as in WSN4QoL, the study involves a real implementation of the proposed architecture in different scenarios, including nursing-house, home, and hospital environments. Besides the security aspects related to the communication of patients' data, the paper shows that the architecture is flexible enough to allow for querying the fixed wireless sensor network nodes and the mobile BANs over heterogeneous communication technologies. Authors also stressed that a tighter integration between the wireless sensor network and the wearable systems can be achieved through the use of 6LoWPAN technologies and that biomedical sensor positioning would be crucial to detect the location of people at any place and any time [47].

In this line the work of [10, 48–50], where the paradigm of the Internet of Things, that is, 6LoWPAN, is applied to healthcare and BSN systems, is worth mentioning. In particular, while [47] focuses on simulation results to assess the performance of a proposed distributed handover procedure to support body sensors mobility and continuous access, in the other papers, preliminary implementations approaches in TinyOS [21] are illustrated with particular emphasis on handling mobility and inter-BSNs communications.

Achieving location of the patients is one of the goals of the work presented in [10] and sketched in Figure 13. In this paper, however, it is supposed that a wearable system built upon smart shirts, such as, [34], is completely hardware-independent from the positioning subsystem built upon devices that each patient is also supposed to carry. Both subsystems communicate with an IEEE 802.15.4-enabled wireless sensor network acting as distribution network between the terminals (patients) and the gateway to the public Internet-based network. Once again, however, the two subsystems (sensor measurements reporting and patients' localization) are independent of each other, while in WSN4QoL the goal is to make them converge into a single network able to support different services simultaneously. The prototype has been implemented and tested in a real-world scenario within hospital facilities by equipping up to ten patients with this equipment, and positive feedback was received by the hospital personnel, paving the way to future applications.

The two-tier architecture is also the basis to the work presented in [51, 52], where scalability performance is evaluated considering the IEEE 802.15.4 at the lower tier and WLAN/IEEE 802.11 at the upper tier of a healthcare monitoring system. End-to-end packet delay and packet access time to WLAN have been evaluated as a function of the number of concurrent BANs (up to 50). Authors claim that there is need for choosing carefully network parameters, because the interaction of high-data-rate streams, such as EEG, with lower-rate streams, such as EKG or blood pressure data, causes some unwanted effects on the packet jitter of the latter. Moreover, in [53] an overall comparison between GPRS and WLAN communication technologies for the higher tier of the system architecture is presented. The study clearly shows that GPRS and WLAN have complementary power and delay profiles: GPRS has lower power consumption to keep network connectivity and send data, but delays might be high, whereas WLAN has higher energy cost but lower delays. Finally, the paper [54], where a mapping of the quality of service requirements for e-Health on the QoS classes of 3GPP networks is presented, is worth mentioning.

A different technological approach is pursued in [55, 56]. In this work, to solve the issues related to RF interferences, authors investigate the potentiality of using infrared communication for data transmissions in mobile healthcare contexts. Simulation of scenarios of line of sight and diffuse configurations show that it is theoretically possible to achieve optimal outage probabilities with very low transmission power, which would help improve the energy efficiency of

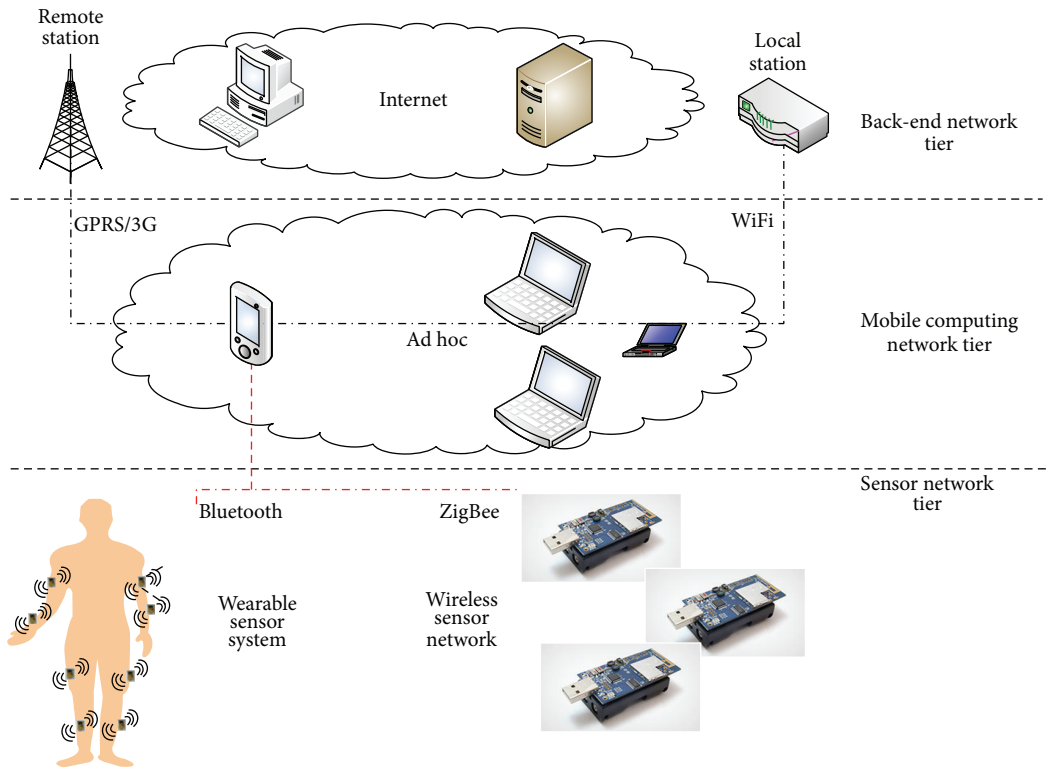


Figure 12: Healthcare system hierarchical network architecture in wireless sensor networks [9].

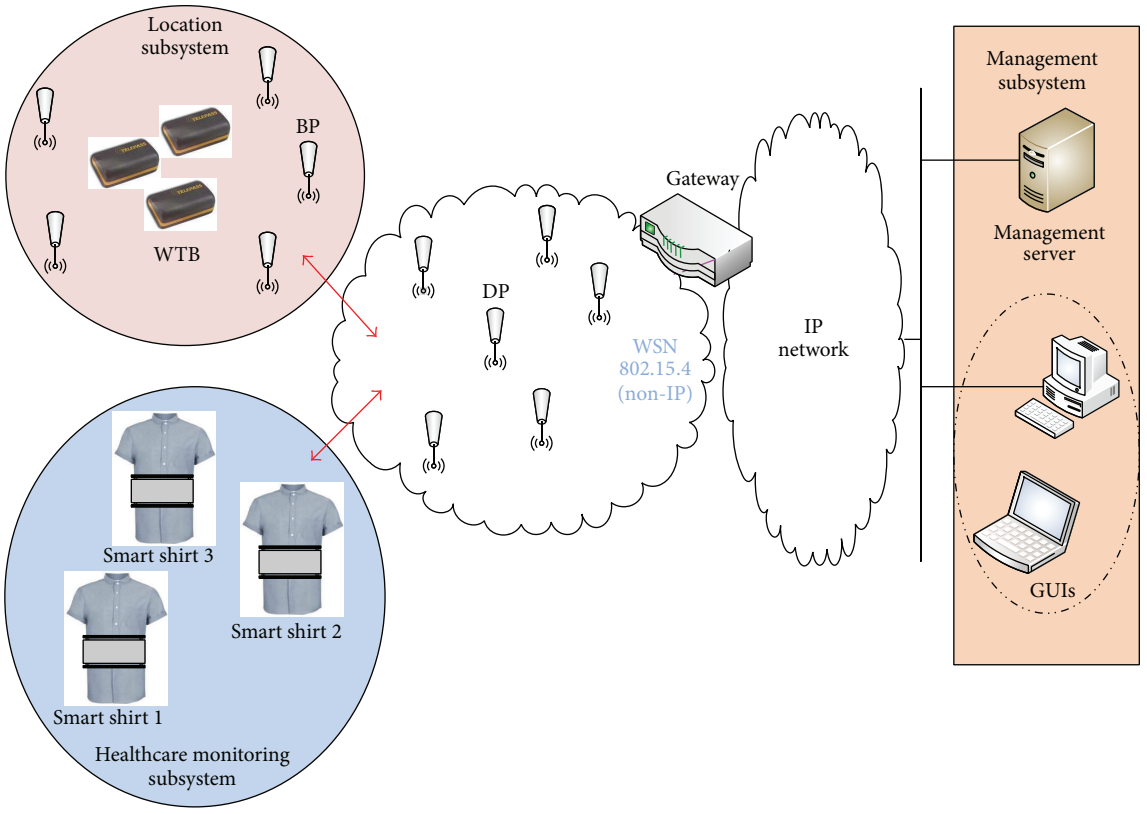


Figure 13: Healthcare monitoring and location systems [10].

such systems. However, although promising, these results need to be confirmed by experimental trials, still far to be achieved.

7. Conclusions

In this paper, the WSN4QoL project has been described with particular emphasis on its challenges and objectives in the area of proposing efficient WSN-based solutions for pervasive healthcare applications. NC techniques and distributed people localization mechanisms have been implemented on real WSN testbeds. Preliminary tests in real working environments gave promising results, in line with our expectations.

Future work will include the implementation of the proposed solutions in real medical devices, as well as repeating these tests on a larger scale testbed. Overall, we believe that the intelligent implementation of the solutions proposed in a self-organized WSN will pave the way for a pervasive healthcare system that is free of economic burdens and is able to focus on the real needs of patients regardless of their age span.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the European Commission under the Marie Curie IAPP WSN4QoL project, Grant no. IAPP-GA-2011-286047. An earlier version of this paper appeared in the Proceedings of the IEEE 15th International Conference on e-Health Networking, Application & Services (HealthCom), 2013 [57].

References

[1] "Midyear population, by age and sex," International database. Table 094.

[2] "Report of the second world assembly on aging," Tech. Rep., United Nations, Madrid, Spain, 2002.

[3] K. Kinsella and V. Velkoff, "An aging world: 2001," series p95/01-1, U.S. Census Bureau, U.S. Government Printing Office, Washington, DC, USA, 2001.

[4] http://www.hhmglobal.com/knowledge-bank/articles/health-spending-projections-through-2015-changes-on-the-horizon.

[5] ETSI, "Machine to Machine Communications (M2M): Use Cases of M2M Applications for eHealth," Draft TR 102732 v0.4.1, 2011.

[6] "Wsn4qol: Wireless sensor networks for quality of life," 2014, http://www.wsn4qol.eu/.

[7] P. Harrop and R. Das, "Wireless sensor networks (wsn) 2012–2022: forecasts, technologies, players—the new market for ubiquitous sensor networks (usn)," Tech. Rep. IDTechEx, 2012.

[8] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2688–2710, 2010.

[9] Y. M. Huang, M. Y. Hsieh, H. C. Chao, S. H. Hung, and J. H. Park, "Pervasive, secure access to a hierarchical sensor-based healthcare monitoring architecture in wireless heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 400–411, 2009.

[10] G. López, V. Custodio, and J. I. Moreno, "LOBIN: E-textile and wireless-sensor-network-based platform for healthcare monitoring in future hospital environments," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 6, pp. 1446–1458, 2010.

[11] B. Otal, L. Alonso, and C. Verikoukis, "Highly reliable energy-saving mac for wireless body sensor networks in healthcare systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 553–565, 2009.

[12] K. Prabh, F. Royo, S. Tennina, and T. Olivares, "Banmac: an opportunistic mac protocol for reliable communications in body area networks," in *Proceedings of the IEEE 8th International Conference on Distributed Computing in Sensor Systems (DCOSS '12)*, pp. 166–175, 2012.

[13] E. Ibarra, A. Antonopoulos, E. Kartsakli, and C. Verikoukis, "Energy harvesting aware hybrid mac protocol for wbans," in *Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications Services (Healthcom '13)*, pp. 120–124, 2013.

[14] Institute of Electrical and Electronics Engineers, "IEEE Std. 802.15.4-2006, IEEE Standard for Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)," Institute of Electrical and Electronics Engineers, New York, NY, USA, 2006.

[15] A. Koubâa, A. Cunha, and M. Alves, "A time division beacon scheduling mechanism for IEEE 802.15.4/zigbee cluster-tree wireless sensor networks," in *Proceedings of the 19th Euromicro Conference on Real-Time Systems (ECRTS '07)*, pp. 125–135, July 2007.

[16] S. Tennina, A. Koubaa, R. Daidone et al., "IEEE 802.15.4 and ZigBee as enabling technologies for low-power wireless systems with quality-of-service constraints," in *Briefs in Electrical and Computer Engineering*, p. 464, Springer, 2013.

[17] T. Ho, M. Médard, R. Koetter et al., "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[18] R. Bassoli, H. Marques, J. Rodriguez, K. Shum, and R. Tafazolli, "Network coding theory: a survey," *IEEE Communications Surveys Tutorials*, vol. 15, pp. 1950–1978, 2013.

[19] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[20] Telosb mote platform, http://www.memsic.com/.

[21] Tinyos, 2012, http://www.tinyos.net/.

[22] J.-H. Hauer, "Tkn15.4: An IEEE 802.15.4 mac—implementation for tinyos 2," Tech. Rep. TKN-08-003, Technical University, Telecommunication Networks Group, Department Telecommunication Networks (TKN), Berlin, Germany, 2009.

[23] J.-H. Hauer, R. Daidone, R. Severino et al., "Poster abstract: an open-source ieee 802.15.4 mac implementation for tinyos 2.1," in *Proceedings of the 8th European Conference on Wireless Sensor Networks (EWSN '11)*, Bonn, Germany, 2011.

[24] A. Antonopoulos and C. Verikoukis, "Network-coding-based cooperative ARQ medium access control protocol for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 601321, 9 pages, 2012.

[25] S. Tennina, M. D. Renzo, F. Graziosi, and F. Santucci, "Esd: a novel optimisation algorithm for positioning estimation of wsns in gps-denied environments—from simulation to experimentation," *International Journal of Sensor Networks*, vol. 6, pp. 131–156, 2009.

[26] S. Tennina, M. di Renzo, F. Graziosi, and F. Santucci, "Locating zigbee nodes using the tis cc2431 location engine: a testbed platform and new solutions for positioning estimation of wsns in dynamic indoor environments," in *Proceedings of the 1st ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments (MELT '08)*, pp. 37–42, New York, NY, USA, September 2008.

[27] Chipcon, "C2430dk development kit datasheet," 2009, http://www.ti.com/.

[28] "Winprop software package," 2013, http://www.awe-communications.com/.

[29] M. Conti, S. K. Das, C. Bisdikian et al., "Looking ahead in pervasive computing: challenges and opportunities in the era of cyberphysical convergence," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 2–21, 2012.

[30] K.-D. Kim and P. R. Kumar, "Cyber-physical systems: a perspective at the centennial," *Proceedings of the IEEE*, vol. 100, pp. 1287–1308, 2012.

[31] H.-H. Ku and C.-M. Huang, "Web2OHS: a Web2.0-based omnibearing homecare system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 224–233, 2010.

[32] H. Jumaa, P. Rubel, and J. Fayn, "An XML-based framework for automating data exchange in healthcare," in *Proceedings of the 12th IEEE International Conference on e-Health Networking, Application and Services (Healthcom '10)*, pp. 264–269, July 2010.

[33] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, pp. 1–17, 2012.

[34] M. di Rienzo, P. Meriggi, F. Rizzo et al., "Textile technology for the vital signs monitoring in telemedicine and extreme environments," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 711–717, 2010.

[35] XSENS, "Xsens research," 2012, http://www.xsens.com/technology/research/.

[36] L. Lhotska, O. Stepankova, M. Pechoucek, B. Simak, and J. Chod, "ICT and eHealth projects," in *Proceedings of the Technical Symposium at ITU Telecom World (ITU WT '11)*, pp. 57–62, October 2011.

[37] T. Yilmaz, R. Foster, and Y. Hao, "Detecting vital signs with wearablewireless sensors," *Sensors*, vol. 10, no. 12, pp. 10837–10862, 2010.

[38] G. J. Mandellos, M. N. Koukias, I. S. Styliadis, and D. K. Lymberopoulos, "E-SCP-ECG + protocol: an expansion on SCP-ECG protocol for health telemonitoringpilot implementation," *International Journal of Telemedicine and Applications*, vol. 2010, Article ID 137201, 17 pages, 2010.

[39] H. Wang, D. Peng, W. Wang, H. Sharif, H.-H. Chen, and A. Khoynezhad, "Resource-aware secure ECG healthcare monitoring through body sensor networks," *IEEE Wireless Communications*, vol. 17, no. 1, pp. 12–19, 2010.

[40] M. Xu, A. Goldfain, A. R. Chowdhury, and J. Dellostritto, "Towards accelerometry based static posture identification," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '2011)*, pp. 29–33, 2011.

[41] "Rd3965mma7660fc: Zstar3 featuring the mma7660fc," 2012, http://www.freescale.com/webapp/sps/site/prod_summary.jsp?code=RD3965MMA7660FC.

[42] S. Abbate, M. Avvenuti, G. Cola, P. Corsini, J. Light, and A. Vecchio, "Recognition of false alarms in fall detection systems," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '2011)*, pp. 23–28, January 2011.

[43] "Shimmer: Wireless sensor solutions," 2012, http://www.shimmer-research.com/.

[44] R. Rednic, J. Kemp, E. Gaura, and J. Brusey, "Networked body sensing: enabling real-time decisions in health and defence applications," in *Proceedings of the International Conference on Advanced Computer Science and Information Systems (ICACSIS '11)*, pp. 17–23, December 2011.

[45] R. L. Dellaca, A. Gobbi, L. Govoni, D. Navajas, A. Pedotti, and R. Farré, "A novel simple internet-based system for real time monitoring and optimizing home mechanical ventilation," in *Proceedings of the International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED '09)*, pp. 209–215, February 2009.

[46] "Acme systems," 2012, http://www.acmesystems.it/.

[47] J. Caldeira, J. Rodrigues, and P. Lorenz, "Toward ubiquitous mobility solutions for body sensor networks on healthcare," *IEEE Communications Magazine*, vol. 50, pp. 108–115, 2012.

[48] D. Singh, U. S. Tiwary, and W.-Y. Chung, "IP-based ubiquitous healthcare system," in *Proceedings of the International Conference on Control, Automation and Systems (ICCAS '08)*, pp. 131–136, October 2008.

[49] D. Singh, H.-P. Kew, U. S. Tiwary, H.-J. Lee, and W.-Y. Chung, "Global patient monitoring system using IP-enabled ubiquitous sensor network," in *Proceedings of the WRI World Congress on Computer Science and Information Engineering (CSIE '09)*, pp. 524–528, April 2009.

[50] A. J. Jara, M. A. Zamora, and A. F. G. Skarmeta, "An architecture based on internet of things to support mobility and security in medical environments," in *Proceedings of the 7th IEEE Consumer Communications and Networking Conference (CCNC '10)*, pp. 1–5, January 2010.

[51] J. Misic and V. B. Misic, "Bridge performance in a multitier wireless network for healthcare monitoring," *IEEE Wireless Communications*, vol. 17, no. 1, pp. 90–95, 2010.

[52] J. Misic and V. Misic, "Bridging between ieee 802.15.4 and IEEE 802.11b networks for multiparameter healthcare sensing," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 435–449, 2009.

[53] K. Wac, M. Bargh, B.-J. van Beijnum, R. Bults, P. Pawar, and A. Peddemors, "Power- and delay-awareness of health telemonitoring services: the mobihealth system case study," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 525–536, 2009.

[54] L. Skorin-Kapov and M. Matijasevic, "Analysis of QoS requirements for e-Health services and mapping to evolved packet system QoS classes," *International Journal of Telemedicine and Applications*, vol. 2010, Article ID 628086, 18 pages, 2010.

[55] S. S. Torkestani, N. Barbot, S. Sahuguede, A. Julien-Vergonjanne, and J. P. Cances, "Performance and transmission power bound analysis for optical wireless based mobile

healthcare applications," in *Proceedings of the IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '11)*, pp. 2198–2202, September 2011.

[56] S. S. Torkestani, A. Julien-Vergonjanne, and J. P. Cances, "Mobile healthcare monitoring in hospital based on diffuse optical wireless technology," in *Proceedings of the IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC '10)*, pp. 1055–1059, September 2010.

[57] S. Tennina, E. Kartsakli, and A. Lalos, "Wsn4qol: wireless sensor networks for quality of life," in *Proceedings of the IEEE 15th International Conference on e-Health Networking, Application & Services (HealthCom '13)*, Lisbon, Portugal, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 347169, 10 pages
http://dx.doi.org/10.1155/2014/347169

Research Article

An Improved User Authentication Protocol for Healthcare Services via Wireless Medical Sensor Networks

Muhammad Khurram Khan[1] and Saru Kumari[2]

[1] *King Saud University, P.O. Box 92144, Riyadh 11653, Saudi Arabia*
[2] *Department of Mathematics, Agra College, Agra, Dr. B. R. A. University, Agra, Uttar Pradesh, India*

Correspondence should be addressed to Muhammad Khurram Khan; mkhurram@ksu.edu.sa

Received 15 November 2013; Accepted 23 December 2013; Published 27 April 2014

Academic Editor: Christos Verikoukis

Healthcare service sector is one of the major applications of Wireless Sensor Networks (WSNs) acknowledged as Wireless Medical Sensor Network (WMSNs). It deploys tiny medical sensor-nodes (*MS*-nodes) on the body of the patient to sense crucial physiological signs which can be accessed and analyzed by registered medical professionals. Recently, Khan et al. analyzed Kumar et al.'s scheme proposed for healthcare applications using WMSNs and observed that the scheme is susceptible to many security weaknesses if an adversary extracts the information from the lost smart card of some user. The adversary can access patient's physiological data without knowing actual password, can deceive medical professionals by sending fake information about patients, can guess the password of a user from the corresponding smart card, and so forth. Besides, the scheme fails to resist insider attack, lacks user anonymity and the session key shared between the user and the *MS*-node is insecure. To overcome these problems, we propose an improved user authentication scheme for healthcare applications using WMSNs. We show that the scheme is free from the identified weaknesses and excels in performance and efficiency scheme.

1. Introduction

Healthcare sector is witnessing a transition from traditional human-labor-dependent services to technology-based smart services. This changeover is the outcome of Wireless Medical Sensor Networks (WMSNs), a transmission technology employed by medical professionals (like nurses, doctors, etc.) to obtain the information like blood pressure, pulse rate, body temperature, ECG of the patients. This is achieved by deploying tiny *MS*-nodes like blood pressure sensors, pulse oximeter, body temperature sensors, and ECG electrodes on the body of patient. The *MS*-nodes sense physiological information from patient's body and then transmit it to the professionals in a wireless manner. Consequently, it cuts the cost of the human labor required for the purpose and facilitates the health professionals to observe and treat the patients as and when required. But patient's personal medical data may be misused by adversaries like corrupt persons, personal enemies, health insurance professionals, and so forth. Thus, there is need for the security of WMSNs to ensure access to patient's physiological information only to the authorized health professionals. Employing a user authentication scheme is a suitable method to achieve the desired security and establish a secure, efficient, and reliable healthcare environment via WMSNs.

After the development of simple user authentication schemes like [1–6], schemes for Wireless Sensor Networks (WSNs) [7–13] have also attracted a large community of researchers. Some work has also been proposed for healthcare applications using WSNs [14–17]. In 2012, Kumar et al. [18] observed that most of the schemes proposed for WSNs such as [9, 10, 12, 13] fall short to provide security and also require heavy computational load and high communication cost. They proposed a user authentication scheme using WMSNs for healthcare applications and called it an Efficient-Strong Authentication Protocol (E-SAP) [18]. They claimed that their scheme achieves mutual authentication between the user and the *MS*-node and also establishes session key between them. They found their scheme finer than other existing protocols concerning cost,

performance, and security. Subsequently, Khan et al. [19] identified that the scheme of Kumar et al. suffers from many security problems if an adversary extracts the information from the stolen smart card of some user. As a consequence, the scheme is exposed to user impersonation attack and insecure session key generation between user and *MS*-node. They showed that the scheme does not go with the authors' claim as the mutual authentication between user and *MS*-node does not imply properly and an adversary can compute the session key to be established between. They also pointed out password guessing attack, insider attack, and *MS*-node impersonation attack on it. They found that if the identity of any user is revealed, it gives chance to many unauthorized/illegal persons to gain the personal medical data of patients and thereby generates problems for an authorized professional.

We feel that in addition to resist the prevalent threats, a user authentication scheme for WMSN should also provide user anonymity. Therefore, we propose a user anonymous authentication scheme using WMSNs eradicating the identified weaknesses of Kumar et al.'s scheme. We aim to provide perfect mutual authentication and secure session key generation between the active participants of the authentication protocol in the scheme. The rest of the paper is arranged as the description follows for the subsequent sections. Section 2 briefly explains the architecture of WMSN and its benefit in healthcare applications. Kumar et al.'s scheme is reviewed in Section 3. Section 4 gives review of the analysis of Kumar et al.'s scheme by Khan et al. The proposed scheme is illustrated in Section 5 along with its security analysis and performance comparison in presented by Sections 6 and 7, respectively. To end with, Section 8 gives the conclusion of this paper. In this paper, we use professional and user interchangeably.

2. Architecture of WMSN and Its Benefits in Healthcare Services

The architecture of the Wireless Medical Sensor Network is depicted by Figure 1. There are four parties involved in the user authentication protocol employing WMSN as described below:

(i) Users: medical professionals like nurses, doctors, and so forth, looking for physiological data of the patient via WMSN.

(ii) *MS*-nodes: tiny sensors like temperature sensor, pulse oximeter, and so forth, deployed on the body of the patients.

(iii) *GW*-node: a powerful master node which plays the role of the registering authority and acts as an interface between the user and the *MS*-node.

(iv) Patients: they are under vigilance of medical professionals by means of *MS*-nodes for treatment.

First three participants are the active parties of the user authentication scheme. *MS*-nodes are tiny sensor having low processing power, limited computational capabilities, and limited energy and storage capacity [20]. *GW*-node is a powerful node with sufficiently large processing power, computational capabilities, and energy and storage capacity [20]. A user registers itself to the *GW*-node to become a valid user of the system. Whenever a user (medical professionals) wishes to obtain the physiological data of the patient, he transmits request message to the *GW*-node. Afterwards, the *GW*-node verifies the legitimacy of the user, if satisfied then it directs the desired *MS*-node(s) to answer to the user's request.

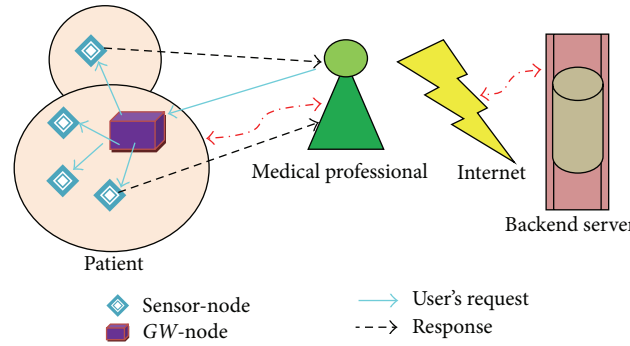


Figure 1: Architecture of WMSN.

Benefits of WMSN in providing healthcare services are as follows:

(i) Improvement in healthcare services,

(ii) Uninterrupted monitoring of patients,

(iii) Saving human labor, time, and money,

(iv) Protecting sensitive and private medical data of the patient from various adversaries.

3. Review of the Scheme Proposed by Kumar et al.

Initially, the *GW*-node chooses three secret keys denoted as J, K and Q, each of 256 bits. The *GW*-node also shares a secret key $K_{gs} = h(Q \parallel ID_g)$ with all deployed *MS*-nodes by means of some key agreement method [21, 22]. The scheme has five phases each of which are described in succession. But before giving detail of each phase of the scheme, we summarize the notations and description used throughout the paper in Notations.

3.1. User Registration Phase. The user (professional) U registers itself to the *GW*-node in registration center of the hospital, in the following manner:

(1) user submits his chosen identity ID_u and password PW_u to *GW*-node using a secure channel;

(2) on receiving $\{ID_u, PW_u\}$, the *GW*-node computes $C_{ug} = E_J(ID_u \parallel ID_g)$ and $N_u = h(ID_u \oplus PW_u \oplus K)$;

(3) *GW*-node stores $\{h(\cdot), C_{ug}, N_u, K\}$ intoa SC and issues $SC = \{h(\cdot), C_{ug}, N_u, K\}$ to U, where K is a long-term secretkey of the *GW*-node.

3.2. Patient Registration Phase. A patient has to register itself in registration center of the hospital [23]. Patient submits her/his name to the registration center. On receiving patient's name, the registration center chooses a suitable medical sensor kit (i.e., *MS*-nodes and *GW*-node) according to the disease of the patient and assigns medical professionals (users). Then the registration center transmits the identity ID_{pt} of the patient along with medical sensors kit information to the assigned professionals/users. Finally, a technician deploys *MS*-node on the body of the patient.

3.3. Login Phase. A professional logs in to the *GW*-node in order to gain patients' medical data via WMSN. The user inserts her/his *SC* into the smart card reader and inputs ID_u and PW_u. Then the *SC* performs the following:

(1) It calculates $N_u{}^* = h(ID_u \oplus PW_u \oplus K)$ and compares $N_u{}^*$ with the stored N_u. For $N_u{}^* = N_u$, the smart card continues further and else ends this session.

(2) It generates a random nonce M to compute $CID_u = E_K(h(ID_u) \parallel M \parallel S_n \parallel C_{ug} \parallel T_u)$.

(3) It sends login request = $\{CID_u, T_u\}$ to the *GW*-node with T_u as the current timestamp.

3.4. Authentication Phase. On receiving the login request $\{CID_u, T_u\}$ from the professional, the *GW*-node verifies the authenticity of U and computes a message to transmit to the desired *MS*-node in the following manner:

(1) It acquires current timestamp $T_g{}'$ and, for $(T_g{}' - T_u) > \Delta T$, discards the login request; otherwise it proceeds further.

(2) It decrypts CID_u as $D_K(CID_u)$ to obtain $\{h(ID_u)^{\$}, M, S_n, C_{ug}$ and $T_u{}^{\$}\}$. Also, it decrypts C_{ug} as $D_J(C_{ug})$ to obtain $\{ID_u{}^*, ID_g{}^*\}$.

(3) It then computes $h(ID_u{}^*)$ and verifies the equivalences $h(ID_u)^* = h(ID_u)^{\$}$, $ID_g{}^* = ID_g$, and $T_u = T_u{}^{\$}$, if all the three equivalences hold, then it believes the login request to come from U; otherwise it terminates the login session.

(4) It acquires T_g, another current timestamp and computes $A_u = E_{SKgs}(ID_u \parallel S_n \parallel M \parallel T_g \parallel T_u)$. Then, the *GW*-node sends $\{A_u, T_g\}$ to the *MS*-node.

When the *MS*-node receives $\{A_u, T_g\}$ from the *GW*-node, it performs the following steps:

(5) It acquires current timestamp $T_s{}'$ and, for $(T_s{}' - T_g) > \Delta T$, discards the received message; otherwise it continues further.

(6) The *MS*-node S_n performs the decryption $D_{SKgs}(A_u)$ and obtains $\{ID_u{}^*, S_n{}^*, M^*, T_g{}^*, T_u\}$ to make sure that the request has come from the legal *GW*-node.

(7) It compares $S_n{}^*$ with S_n and T_g with $T_g{}^*$, and if any of these fail to match, then it discards the message; otherwise it continues further.

(8) It computes session key $Ksess_{U\text{-}Sn} = h(ID_u{}^* \parallel S_n \parallel M^* \parallel T_u)$. Then it acquires T_s, another current timestamp, and computes $L = E_{Ksess_{U\text{-}Sn}}(S_n \parallel M^* \parallel T_s)$. The *MS*-node sends $\{L, T_s\}$ to the user.

When U receives $\{L, T_s\}$ from the *MS*-node, its *SC* performs the following steps:

(9) It acquires current timestamp $T_u{}'$ and, for $(T_u{}' - T_s) > \Delta T$, discards the received message. Or else, it proceeds further.

(10) It computessession key $Ksess_{U-Sn} = h(ID_u \parallel S_n \parallel M \parallel T_u)$. Then it performs the decryption $D_{Ksess_{U\text{-}Sn}}(L)$ and obtains S_n and M^*. It compares $S_n{}^*$ with S_n, and M^* with M; if both the equivalences hold only, then the session key is assumed to be established securely.

3.5. Password Change Phase. U can change her/his password through the following stepwise procedure:

(1) U inserts her/his *SC* into the terminal and inputs ID_u and PW_u.

(2) Then *SC* computes $N_u{}^* = h(ID_u \oplus PW_u \oplus K)$ and compares $N_u{}^*$ with N_u. If both the values match allows the user to enter a new password, otherwise discards the process.

(3) U enters new password $(PW_u)_{new}$.

(4) *SC* computes $(N_u)_{new} = h(ID_u \oplus (PW_u)_{new} \oplus K)$ and then replaces N_u with $(N_u)_{new}$.

4. Review of the Analysis of Kumar et al.'s Scheme

This section presents a review of the security problems of Kumar et al.'s scheme identified by Khan et al. [19]. This analysis is based on the assumption that an adversary U_a can recover [24, 25] the information stored in smart card.

If U_a extracts values $\{h(\cdot), C_{ug}, N_u, K\}$ from the lost *SC* of a user, then he holds the master key K which is stored in the *SC* of each user (professional). Consequently, the scheme becomes vulnerable to different attacks described as follows.

4.1. User Impersonation Attack. Having master key K in hand, U_a can impersonate any user of the system to obtain patient's physiological information. To impersonate U, the attacker U_a intercepts the login request $\{CID_u, T_u\}$ of U and decrypts CID_u as $D_K(CID_u) = (h(ID_u) \parallel M \parallel S_n \parallel C_{ug} \parallel T_u)$ to obtain $\{h(ID_u), M, S_n, C_{ug}\}$. Now U_a acquires a current timestamp T_a and a random nonce M_a. Then computes $CID_a = E_K(h(ID_u) \parallel S_n \parallel C_{ug} \parallel T_a)$ and sends $\{CID_a, T_a\}$ to *GW*-node. Clearly this message will successfully go through *GW*-node authentication test as it contains valid values $\{h(ID_u), S_n, C_{ug}\}$ and fresh values $\{M_a, T_a\}$.

4.2. Lacks User Anonymity. U_a can obtain the hashed value of the identity of any user by decrypting the first component

of the login request. For instance, if U_a intercepts the login request $\{CID_u, T_u\}$ of U, then he can obtain $h(ID_u)$ by decrypting CID_u using K. Having hashed value of user's identity $h(ID_u)$ in hand, U_a can guess the corresponding identity ID_u of U. Thus, the scheme fails to provide user anonymity.

4.3. Password Guessing Attack. We further extend the above two threats to a more harmful vulnerability. If U_a successfully guesses the identity ID_u of the user from whose smart card he extracts the secret key K, then he can guess the password PW_u of U. For this, U_a guesses PW_a as the probable password, computes $N_u{}^* = h(ID_u \oplus PW_a \oplus K)$, and verifies if $N_u{}^* = N_u$. If so, it implies success of U_a in guessing the PW_u of U. In fact, it is complete violation of security since U_a holds user's SC along with user's identity ID_u and password PW_u so he can behave as the legal user U.

4.4. Illegal Logged-In Users Using Legal Identity. U_a can guess the identity ID_u of any user as described in Section 4.2; he can misuse ID_u for crafting other damage to the security of the scheme as described below.

(1) U_a applies for her/his registration by submitting ID_u and PW_a, where PW_a is a random password chosen by U_a.

(2) In return, the GW-node provides U_a a $SC_a = \{h(\cdot), C_{ug}, N_a, K\}$ with $N_a = h(ID_u \oplus PW_a \oplus K)$ and $C_{ug} = E_J(ID_u \parallel ID_g)$.

The role of password in the login-authentication procedure of the scheme is up to confirming the legitimacy of the user by her/his smart card. From then on, only user's identity ID_u is used to authenticate U at the GW-node. As a result, there are two pictures.

(i) U_a can successfully log in as the legal user U with the received $SC_a = \{h(\cdot), C_{ug}, N_a, K\}$. U_a inserts her/his SC into the terminal and inputs ID_u and PW_a. Once ID_u and PW_a are verified, SC_a computes $CID_a = E_K(h(ID_u) \parallel M_a \parallel S_n \parallel C_{ug} \parallel T_a)$ and sends the login request $\{CID_a, T_a\}$ to the GW-node. Clearly, the GW-node considers it as a valid login request from the legitimate user U since it is computed using valid ID_u in C_{ug}.

(ii) U_a has open option to distribute the user's identity ID_u among malicious persons interested to obtain patient's private health data in an illicit way. These persons can register themselves in similar manner as just explained in the previous scenario and can access data through MS-node. U_a can also distribute the values $\{h(ID_u), S_n, C_{ug}\}$ in place of ID_u among these persons. Then it is possible to impersonate U as described in Section 4.1. In case such an illegal access is detected by the system, it will raise a question on the credibility of the valid user (medical professional) whose identity ID_u is misused byU_a.

4.5. Insecure Session-Key. U_a can compute the session key to be used between a user and a MS-node during a particular session. Suppose U_a recovers the values $\{h(ID_u), M, S_n\}$ out of CID_u of the intercepted login request of U. Then he attempts to guess the identity ID_u as described in Section 4.2 and uses timestamp T_u from the corresponding intercepted login request $\{CID_u, T_u\}$. Then U_a can easily compute the session key $Ksess_{U\text{-}Sn} = h(ID_u \parallel S_n \parallel M \parallel T_u)$ to be used by U and the MS-nodewith identity S_n. Hence, the shared session key $Ksess_{U\text{-}Sn}$ is insecure and U_a can decrypt the confidential messages communicated between MS-node and U.

4.6. MS-Node Impersonation Attack. An active attacker U_a having secret key K obtained from a lost or stolen SC can perform decryption of CID_u's for as many users as he wants. As a result, he can obtain the hashed value like $h(ID_u)$ of all the target users. Next, U_a can guess the identity ID_u for each $h(ID_u)$ and tabulates the values $\{h(ID_u), ID_u\}$. After that, U_a can impersonate the MS-node to deceit legitimate users as explained below.

(1) As U_a finds a login request $\{CID_u, T_u\}$ on the network, he intercepts and blocks it and quickly decrypts CID_u to see if $h(ID_u)$ included in it is present in the table maintained or not. If not then it relays the login request to GW-node.

(2) But if $h(ID_u)$ exists in the tabular record, then U_a keeps the login request blocked and uses ID_u from the record, values $\{M, S_n\}$ from current decryption, and T_u from login request; U_a quickly computes $Ksess_{U\text{-}Sn} = h(ID_u \parallel S_n \parallel M \parallel T_u)$.

(3) It computes $L = E_{Ksess_{U\text{-}Sn}}(S_n \parallel M \parallel T_a)$ and sends $\{L, T_a\}$ to U, where T_a is the current timestamp chosen by U_a.

(4) Obviously L will qualify the verification test at the user side as it consists of valid $\{S_n, M\}$ and fresh timestamp T_a.

It is noticeable that $Ksess_{U\text{-}Sn}$, the common session key is computed by U and U_a but U believes it to be confidential between him and the MS-node. Moreover, U_a can misguide the user doctor by sending fake data about the patient. Consequently, the patient may receive false treatment, thus denying the goal of healthcare through WMSN.

4.7. Lacking of Mutual Authentication between (i) MS-Node and GW-Node, (ii) U and MS-Node. In Kumar et al.'s scheme, after verifying the login request of U, GW-node computes and sends an ensuring message $\{A_u, T_g{}'\}$ to the required MS-node. Undoubtedly, the equivalence $S_n{}^* = S_n$ confirms the legality of GW-node to MS-node but reverse is not achieved. Thus, GW-node has no way to ensure itself of connecting with real MS-node. Hence, mutual authentication between MS-node and GW-node is not achieved in the scheme.

Besides, the authors claim that their scheme achieves mutual authentication between MS-node and user U. Mutual authentication between U and MS-node is established using the session key $Ksess_{U\text{-}Sn} = h(ID \parallel S_n \parallel M \parallel T_u)$.

But as shown in Sections 4.5 and 4.6, U_a can compute $Ksess_{U\text{-}Sn}$ and impersonate S_n, respectively. Therefore, mutual authentication between U and MS-node is not achieved in the scheme.

4.8. Insider Attack. For convenience people use the same password for more than one application. During registration phase of the scheme, user submits her/his password plaintext PW_u to GW-node. So, the system administrator at the GW-node easily comes to know the password of each user and he can use it to impersonate U at servers, where U is registered with the same password. Although authors assume the hospital registration center as a trusted authority, we think that often the trustworthy breaches the trust. Therefore, plaintext password PW_u should not be submitted to any second party.

5. The Proposed Scheme

The proposed scheme has the same number of phases as in Kumar et al.'s scheme. Each of the phases is detailed below along with Tables 1, 2, and 3. The GW-nodekeeps only one master secret key K (length 256 bits). Besides, the GW-node shares a secret key $K_{gs} = h(K \parallel ID_g)$ with MS-nodes using some key agreement method [21, 22].

5.1. User Registration Phase. The user (professional) U registers itself to the GW-node in registration center of the hospital, in the following manner.

(1) User choses her/his identity ID_u and submits it to the GW-node using a secure channel.

(2) On receiving ID_u the GW-node computes $C_{ug} = E_K(ID_u \parallel ID_g)$, $K_u = h(K \parallel ID_u \parallel ID_g)$, and $K_g = h(ID_g \parallel K)$.

(3) GW-node stores $\{h(\cdot), C_{ug}\}$ into a SC and provides $SC = \{h(\cdot), C_{ug}\}$ along with values $\{K_u, K_g\}$ to U through the secure channel.

(4) On obtaining $SC = \{h(\cdot), C_{ug}\}$ and $\{K_u, K_g\}$, the user U chooses his password PW_u and computes $N_u = h(ID_u \parallel PW_u \parallel K_u)$, $PK_u = K_u \oplus (ID_u \parallel PW_u)$, and $PK_g = K_g \oplus (PW_u \parallel ID_u)$. Finally, U inserts N_u, PK_u, and PK_g in SC, so that $SC = \{h(\cdot), C_{ug}, N_u, PK_u, PK_g\}$.

5.2. Patient Registration Phase. This phase is identical to that in Kumar et al.'s scheme so we avoid its explanation here.

5.3. Login Phase. A professional logs in the GW-node in order to gain patients' medical data via WMSN. The user inserts her/his SC into the smart card reader and inputs ID_u and PW_u. Then the SC performs the following.

(1) It retrieves $K_u = P_{Ku} \oplus (ID_u \parallel PW_u)$, $K_g = PK_g \oplus (PW_u \parallel ID_u)$ and computes $N_u^* = h(ID_u \parallel PW_u \parallel K_u)$. For $N_u^* = N_u$ it continues further; otherwise it stops the session.

(2) It generates a random nonce M and computes $C_{u1} = C_{ug} \oplus h(K_g)$ and $CID_u = E_{Ku}(h(ID_u) \parallel M \parallel S_n \parallel C_{ug} \parallel T_u)$.

(3) SC sends $\{CID_u, C_{u1}, T_u\}$ as login request to GW-node, where T_u is the current timestamp.

5.4. Authentication Phase. When the login request $= \{CID_u, C_{u1}, T_u\}$ from U is received by the GW-node, it executes the following steps.

(1) It acquires current timestamp T_g' and, for $(T_g' - T_u) > \Delta T$, discards the login request; otherwise it proceeds further.

(2) It retrieves $C_{ug} = C_{u1} \oplus h(K_g)$ and decrypts CID_u as $D_K(CID_u)$ to obtain $\{h(ID_u)^{\S}, M, S_n, C_{ug}{}^{\S}$ and $T_u{}^{\S}\}$. It verifies the equivalence $C_{ug}{}^{\S} = C_{ug}$, and if correct, then it decrypts C_{ug} as $D_K(C_{ug})$ to obtain $ID_u{}^*$ and $ID_g{}^*$.

(3) It then computes $h(ID_u{}^*)$ and verifies the equivalences $h(ID_u)^* = h(ID_u)^{\S}$, $ID_g{}^* = ID_g$, and $T_u = T_u{}^{\S}$, if all the three equivalences hold then believes the login request to come from U; otherwise terminates the login session.

(4) It acquires T_g as current timestamp, computes $C_{g1} = K_g \oplus (M \parallel T_g \parallel T_u)$, and sends $\{C_{g1}, T_g\}$ to U. It acquires T_{gs} as another current timestamp and computes $C_{g2} = h(K_{gs}) \oplus (CID_u \parallel T_g \parallel M \parallel T_u)$ and $A_u = h(CID_u \parallel K_{gs} \parallel T_g \parallel S_n \parallel T_{gs})$. Then, the GW-node sends $\{C_{g2}, A_u, T_{gs}\}$ to the MS-node.

On receiving $\{C_{g1}, T_g\}$ from the GW-node, U verifies the legitimacy of GW-node as follows.

(5) It checks if $(T_u' - T_g) > \Delta T$; if so, it dumps the session; otherwise it continues further.

(6) It obtains $(M^* \parallel T_g{}^* \parallel T_u{}^*) = C_{g1} \oplus K_g$ and verifies the equivalence $M^* = M$, $T_g{}^* = T_g$, and $T_u{}^* = T_u$; if each holds, then GW-node is authenticated; otherwise it terminates the login session.

After this mutual authentication, U and GW-node compute $Ksess_{U\text{-}GW} = h(M \parallel ID_u \parallel T_g)$, as the session key.

On receiving $\{C_{g2}, A_u, T_{gs}\}$ from the GW-node, the MS-node performs the following operations.

(7) It checks if $(T_s' - T_{gs}) > \Delta T$; if so, terminates the session; otherwise it proceeds further.

(8) It obtains $(CID_u{}^* \parallel T_g{}^* \parallel M^* \parallel T_u{}^*) = C_{g2} \oplus h(K_{gs})$, computes $A_u{}^* = h(CID_u{}^* \parallel K_{gs} \parallel T_g{}^* \parallel S_n \parallel T_{gs})$, and compares $A_u{}^*$ with A_u. The equivalence $A_u{}^* = A_u$ verifies the legitimacy of the GW-node and hence of U.

(9) It acquires T_{sg} as current timestamp, computes $C_{s1} = h(T_g \parallel K_{gs} \parallel T_{sg}) \oplus h(CID_u \parallel S_n)$, and sends $\{C_{s1}, T_{sg}\}$ to the GW-node. Also it computes $C_{s2}{}^* = h(S_n \parallel M^* \parallel T_u{}^* \parallel Ts)$, where T_s is another

Table 1: User registration phase of the Proposed scheme.

User (U)		GW-node
User registration phase:		
Chooses ID_u		$C_{ug} = E_K(ID_u \parallel ID_g)$
	$\xrightarrow{\{ID_u\}}$	$K_u = h\left(K \parallel ID_u \parallel ID_g\right)$
	$\xleftarrow{SC = \{h(\cdot), C_{ug}\}\ \&\ \{K_u, K_g\}}$	$K_g = h\left(ID_g \parallel K\right)$
Chooses PW_u		
$N_u = h(ID_u \parallel PW_u \parallel K_u)$		
$PK_u = K_u \oplus (ID_u \parallel PW_u)$		
$PK_g = K_g \oplus (PW_u \parallel ID_u)$		
Inserts N_u, PK_u, and PK_g in SC so that $SC = \{h(\cdot), C_{ug}, N_u, PK_u, PK_g\}$		

Table 2: Login and authentication phase of the Proposed scheme.

User (U)	GW-node	MS-node
Login and authentication phase:		
U: inserts ID_u and PW_u		
SC: $K_u \leftarrow PK_u \oplus (ID_u \parallel PW_u)$,		
$K_g \leftarrow PK_g \oplus (PW_u \parallel ID_u)$,		
$N_u^* = h(ID_u \parallel PW_u \parallel K_u)$.		
For $N_u^* = N_u$		
$C_{u1} = C_{ug} \oplus h(K_g)$,		
$CID_u = E_{Ku}\left(h(ID_u) \parallel M \parallel S_n \parallel C_{ug} \parallel T_u\right)$.		
$\xrightarrow{\{CID_u, Cu_1, T_u\}}$	For $(T_g' - T_u) \le \Delta T$	
	$C_{ug} \leftarrow C_{u1} \oplus h(K_g)$,	
	$\{h(ID_u)^{\S}, M, S_n, C_{ug}{}^{\S}, T_u{}^{\S}\} \leftarrow D_K(CID_u)$	
	For $C_{ug}{}^{\S} = C_{ug}$	
	$\{ID_u{}^*, ID_g{}^*\} \leftarrow D_K(C_{ug})$	
	For $h(ID_u{}^*) = h(ID_u)^{\S}$, $ID_g{}^* = ID_g$ & $T_u = T_u{}^{\S}$	
	$C_{g1} = K_g \oplus (M \parallel T_g \parallel T_u)$	
For $\left(T_u' - T_g\right) \le \Delta T$ $\xleftarrow{\{C_{g1}, T_g\}}$	$C_{g2} = h(K_{gs}) \oplus \left(CID_u \parallel T_g \parallel M \parallel T_u\right)$	
$\left(M^* \parallel T_g{}^* \parallel T_u{}^*\right) \leftarrow C_{g1} \oplus K_g$	$A_u = h\left(CID_u \parallel K_{gs} \parallel T_g \parallel S_n \parallel T_{gs}\right)$	
For $M^* = M, T_g{}^* = T_g$ and $T_u{}^* = T_u$	$\xrightarrow{\{C_{g2}, A_u, T_{gs}\}}$	
	$Ksess_{U-GW} = h\left(M \parallel ID_u \parallel T_g\right)$	For $(T_s' - T_{gs}) \le \Delta T$
		$\left(CID_u{}^* \parallel T_g{}^* \parallel M^* \parallel T_u{}^*\right) \leftarrow C_{g2} \oplus h(K_{gs})$
		$A_u{}^* = h(CID_u{}^* \parallel K_{gs} \parallel T_g{}^* \parallel S_n \parallel T_{gs})$
		For $A_u{}^* = A_u$
	$\xleftarrow{\{C_{s1}, T_{sg}\}}$	$C_{s1} = h\left(T_g \parallel K_{gs} \parallel T_{sg}\right) \oplus h\left(CID_u \parallel S_n\right)$
		$C_{s2}{}^* = h\left(S_n \parallel M^* \parallel T_u{}^* \parallel T_s\right)$
	$\xleftarrow{\{C_{s2}^*, T_s\}}$	
	For $(T_g'' - T_{sg}) \le \Delta T$	
	$\left(h\left(CID_u \parallel S_n\right)\right)^* \leftarrow C_{s1} \oplus h\left(T_g \parallel K_{gs} \parallel T_{sg}\right)$	
	For $\left(h\left(CID_u \parallel S_n\right)\right)^* = h(CID_u \parallel S_n)$	
		$Ksess_{GW-Sn} = h(K_{gs} \parallel T_{sg} \parallel M)$
For $(T_u'' - T_s) \le \Delta T$		
$C_{s2} = h\left(S_n \parallel M \parallel T_u \parallel T_s\right)$		
$C_{s2} = C_{s2}{}^*$		
$\longleftarrow$	$Ksess_{U-Sn} = h\left(M \parallel T_s \parallel S_n\right)$	$\longrightarrow$

Table 3: Password change phase of the Proposed scheme.

User (U)		Smart card (SC)
Password change phase: U: inserts ID_u and PW_u		
	$\xrightarrow{\{ID_u, PW_u\}}$	SC: $K_u \leftarrow PK_u \oplus (ID_u \parallel PW_u)$, $K_g \leftarrow PK_g \oplus (PW_u \parallel ID_u)$, $N_u^* = h(ID_u \parallel PW_u \parallel K_u)$. For $N_u^* = N_u$
	$\xrightarrow{(PW_u)_{new}}$	$(N_u)_{\text{new}} = h(ID_u \parallel (PW_u)_{\text{new}} \parallel K_u)$, $(PK_u)_{\text{new}} = K_u \oplus (ID_u \parallel (PW_u)_{\text{new}})$ and $\left(PK_g\right)_{\text{new}} = K_g \oplus ((PW_u)_{\text{new}} \parallel ID_u)$. $(N_u)_{\text{new}} \leftarrow N_u, (PK_u)_{\text{new}} \leftarrow PK_u$ and $\left(PK_g\right)_{\text{new}} \leftarrow PK_g$

current timestamp of MS-node. Then, the MS-node sends $\{C_{s2}^*, T_s\}$ to U.

On receiving $\{C_{s1}, T_{sg}\}$ from the MS-node, the GW-node performs the following operations.

(10) It checks if $(T_g'' - T_{sg}) > \Delta T$; if so, terminates the session; otherwise it proceeds further.

(11) It obtains $(h(CID_u \parallel S_n))^* = C_{s1} \oplus h(T_g \parallel K_{gs} \parallel T_{sg})$, computes $h(CID_u \parallel S_n)$, and compares it with $(h(CID_u \parallel S_n))^*$. The equivalence $(h(CID_u \parallel S_n))^* = h(CID_u \parallel S_n)$ verifies the legitimacy of MS-node.

After this mutual authentication, GW-node and MS-node compute $Ksess_{GW\text{-}Sn} = h(K_{gs} \parallel T_{sg} \parallel M)$, as the session key.

On receiving $\{C_{s2}^*, T_s\}$ from the MS-node, U performs the following.

(12) It checks if $(T_u'' - T_s) > \Delta T$; if so, it dumps the session, otherwise it proceeds further.

(13) It computes $C_{s2} = h(S_n \parallel M \parallel T_u \parallel T_s)$ and compares it with C_{s2}^*, and for $C_{s2} = C_{s2}^*$ the authenticity of MS-node is verified.

After this mutual authentication, U and MS-node compute $Ksess_{U\text{-}Sn} = h(M \parallel T_s \parallel S_n)$, as the session key.

5.5. Password Change Phase. U can change her/his password in the following manner. For this, U inserts her/his SC into the terminal, inputs her ID_u and PW_u, and opts to change his password. Then the following steps are performed to update a new password.

(1) SC retrieves $K_u = PK_u \oplus (ID_u \parallel PW_u)$, $K_g = PK_g \oplus (PW_u \parallel ID_u)$ and computes $N_u^* = h(ID_u \parallel PW_u \parallel K_u)$. If $N_u^* = N_u$, then it proceeds further after asking for new password; otherwise it discards the password change request.

(2) U enters new password $(PW_u)_{new}$.

(3) SC computes $(N_u)_{new} = h(ID_u \parallel (PW_u)_{new} \parallel K_u)$, $(PK_u)_{new} = K_u \oplus (ID_u \parallel (PK_u)_{new})$, and $(PK_g)_{new} = K_g \oplus ((PW_u)_{new} \parallel ID_u)$.

(4) SC replaces N_u, PK_u, and PK_g with $(N_u)_{new}$, $(PK_u)_{new}$, and$(PK_g)_{new}$, respectively.

6. Analysis of the Security of the Proposed Scheme

This section, examines the security of the proposed scheme. We will display that the proposed scheme is secure under the same assumption subject to which Kumar et al.'s scheme is attackable. The assumption is that an attacker U_a can extract [24, 25] the information stored inside smart card.

6.1. Resisting User Impersonation Attack. To impersonate as the user, U_a has to compute a valid login request. Suppose U_a obtains the lost smart card of U and extracts the values $\{C_{ug}, N_u, PK_u, PK_g\}$ stored in it. Though C_{ug} is involved in both the components $\{CID_u$ and $C_{u1}\}$ of the login request, but without K_g, $h(ID_u)$, and S_n computation of these components is incomplete. To recover K_g from PK_g, the attacker U_a needs to know of user's identity and password. On the contrary, to obtain ID_u from PK_u or PK_g, the attacker U_a should hold K_u or K_g, respectively. Further, it is not feasible to obtain ID_u or K_u from N_u due to noninvertible nature of hash function. Thus, the scheme resists user impersonation attack.

6.2. Providing User Anonymity. If U_a intercepts the login request $\{CID_u, C_{u1}, T_u\}$ of U, then he needs K_u to obtain $h(ID_u)$ by decrypting CID_u. But U_a neither knows K_u nor can recoverit by extracting information $\{h(\cdot), C_{ug}, N_u, PK_u, PK_g\}$ from the lost smart card of some user; say U. To take out K_u from PK_u, the attacker U_a should know user's identity and password. In fact, key K_u required to encrypt/decrypt CID_u is not stored directly in user's smart card and is different for each user. Therefore, U_a cannot obtain $h(ID_u)$ and guess the identity as in Kumar et al.'s scheme. On the other hand, to procure identity ID_u from N_u, PK_u, or PK_g is infeasible. It requires knowledge of keys K_u and K_g to gain ID_u out of PK_u or PK_g, respectively. Moreover, one-way property of hash function does not allow extraction of ID_u out of N_u. Therefore, U_a cannot gain the identity of a user and hence the scheme provides user anonymity.

6.3. Resisting Password Guessing Attack. In order to guess U's password PW_u from $N_u = h(ID_u \parallel PW_u \parallel K_u)$ obtained from the lost SC of U, the attacker U_a requires knowledge of

ID_u and K_u. As described in Section 6.2, U_a cannot gain the identity of a user either from the lost smart card of a user or from an intercepted login request. Besides, K_u is not available as plaintext in U's SC and is not obtainable from PK_u without having exact values of ID_u and PW_u. Thus, the scheme resists password guessing attack.

6.4. Resisting Illegal Logged-In Users Using Legal Identity. Since it is not possible to guess or know the identity ID_u of a logging user, U_a cannot register itself to the GW-node with legal identity ID_u and fake password PW_a. Hence U_a cannot harm the security of the scheme by misusing the identity. As a result, the scenario of many illegal users logged in with legal identity ID_u of a registered user is not possible in the proposed scheme.

6.5. Providing Secure Session Key between Every Pair of the Participating Entities. The proposed scheme establishes session key between every pair of participating entities. Session key between U and GW-node is $Ksess_{U\text{-}GW} = h(M \parallel ID_u \parallel T_g)$ which depends on three values M, ID_u, and T_g. Although user's identity ID_u is fixed, M and T_g are different for each session imparting dynamic nature to $Ksess_{U\text{-}GW}$. However T_g is available in $\{C_{g1}, T_g\}$ from the open network but U_a cannot compute $Ksess_{U\text{-}GW}$ without having M and ID_u. Session key between GW-node and MS-node is $Ksess_{GW\text{-}Sn} = h(K_{gs} \parallel T_{sg} \parallel M)$ which is dynamic because of fresh timestamp T_{sg} and one time usable random number M. Although K_{gs} is fixed but is known only to the GW-node and the MS-node so no one except these two entities can compute the valid $Ksess_{GW\text{-}Sn}$. Moreover, U_a cannot procure M from CID_u without knowing K_u; from C_{g1} without knowing K_{gs}; and from ${C_{s2}}^*$ due to noninvertible nature of hash function. Session key between U and MS-node is $Ksess_{U\text{-}Sn} = h(M \parallel T_s \parallel S_n)$ which an attacker cannot compute without knowing M. Thus, the scheme establishes independent and secure session keys between every pair of the participating entities.

6.6. Resisting Sensor-Node Impersonation Attack. In order to impersonate the MS-node, U_a should be able to compute the response messages sent by it to the GW-node and U. To compute $C_{s1} = h(T_g \parallel K_{gs} \parallel T_{sg}) \oplus h(CID_u \parallel S_n)$ and ${C_{s2}}^* = h(S_n \parallel M^* \parallel {T_u}^* \parallel T_s)$ the knowledge of K_{gs} and M^* is required, respectively. Since $K_{gs} = h(K \parallel ID_g)$ is shared secretly by GW-node with MS-node using some key agreement method [21, 22] and its computation involves master secret key K and identity ID_g of the GW-node, U_a cannot access or compute K_{gs}. Further M/M^* is not retrievable from CID_u, C_{g1}, and C_{g2} without knowing K_u and K_g, respectively. Moreover, one-way property of hash function prohibits extraction of M/M^* from ${C_{s2}}^*$. Hence U_a cannot impersonate the MS-node to make fool of the user and GW-node.

6.7. Providing Mutual Authentication between Every Pair of the Participating Entities. At each of the three ends, any received message undergoes at least two-step verification test to verify the authenticity of the sender. For every message, firstly timestamp is checked for freshness followed by one or more equivalences holding tests. The proposed scheme achieves mutual authentication between U and GW-node by exchange of messages $\{CID_u, C_{u1}, T_u\}$ and $\{C_{g1}, T_g\}$. When GW-node receives $\{CID_u, C_{u1}, T_u\}$ from U, in addition to timestamp freshness test, the equivalences $h({ID_u}^*) = h(ID_u)^{\$}$, ${ID_g}^* = ID_g$, and ${T_u} = {T_u}^{\$}$ are required to guarantee the legitimacy of U. Similarly, for $\{C_{g1}, T_g\}$ received by U from GW-node, the equivalence $M^* = M$, ${T_g}^* = T_g$, and ${T_u}^* = T_u$ should hold to prove the validity of the GW-node.

Mutual authentication between the GW-node and the MS-node is achieved through the messages $\{C_{g2}, A_u, T_{gs}\}$ and $\{C_{s1}, T_{sg}\}$. Corresponding to the message $\{C_{g2}, A_u, T_{gs}\}$, the equivalence ${A_u}^* = A_u$ is imperative to confirm the legitimacy of GW-node and hence of U to MS-node. On the other hand, only the designated MS-node can compute $C_{s1} = h(T_g \parallel K_{gs} \parallel T_{sg}) \oplus h(CID_u \parallel S_n)$ and the authorized GW-node can retrieve correct $h(CID_u \parallel S_n)$ from C_{s1} as the computation and retrieval involves use of K_{gs} hence mutually authenticate the entities to each other.

As just discussed, U is authenticated to the MS-node via message $\{C_{g2}, A_u, T_{gs}\}$ with which GW-node is verified. Finally, the legitimacy of MS-node is ensured to U by means of the equivalence ${C_{s2}} = {C_{s2}}^*$. In this way, our scheme provides perfect mutual authentication.

6.8. Resisting Insider Attack. During registration phase, U submits only his identity ID_u to the GW-node at the hospital registration center. The GW-node provides secret keys K_u and K_g to the user. Then using his chosen password PW_u and identity ID_u, the user U itself computes $N_u = h(ID_u \parallel PW_u \parallel K_u)$ and embeds K_u and K_g as $PK_u = K_u \oplus (ID_u \parallel PW_u)$ and $PK_g = K_g \oplus (PW_u \parallel ID_u)$, respectively. Finally, U inserts N_u, PK_u, and PK_g in SC. Since the insider of the system never receives user's password, privileged insider attack is not applicable on the scheme.

7. Performance Analysis of the Proposed Scheme *via* Comparison

Now, we compare our scheme with Kumar et al.'s scheme [18] to present a comparative analysis of its performance and efficiency. Table 4 is about memory space required by smart card and computational complexity/cost in both the schemes. Table 5 exhibits the performance of both the schemes. For convenience, we assume that the identity ID_u, password PW_u, random numbers $\{M\}$, timestamps $\{T_u, T_g$, etc.$\}$, and outputs of one-way hash function $\{h(ID_u \parallel PW_u \parallel K_u)$, etc.$\}$ are 128-bit long.

Table 4 shows that the memory space required by the smart card in Kumar et al.'s scheme and the proposed scheme is 512 bits and 640 bits, respectively. Further, it is noticeable that our scheme adds some hash functions ($h(\cdot)$) but remarkably cuts the number of time consuming symmetric cryptography operation (S_y) at each of the three ends. The most important aspect is that there is no symmetric

Table 4: Comparison of efficiency: memory space and computational cost/complexity.

	Schemes	
	Kumar et al.'s [18]	Ours
Memory space needed by *SC*	512 bits	640 bits
Computational complexity		
Registration phase (U)	Nil	$1h(\cdot)$
Registration phase (GW-node)	$1h(\cdot) + 1S_y$	$2h(\cdot) + 1S_y$
Login and authentication phase (SC)	$4h(\cdot) + 2S_y$	$6h(\cdot) + 1S_y$
Login and authentication phase (GW-node)	$1h(\cdot) + 3S_y$	$7h(\cdot) + 1S_y$
Login-authentication phase (MS-node)	$1h(\cdot) + 2S_y$	$7h(\cdot)$

Table 5: Comparison of performance.

Security characteristics	Schemes	
	Kumar et al.'s [18]	Ours
Resists user impersonation attack	No	Yes
Resists password guessing attack	No	Yes
Resists multi-logged-in users attack	No	Yes
Resists sensor-node impersonation attack	No	Yes
Resists insider attack	No	Yes
Provides user anonymity	No	Yes
Provides verification mechanism in SC	Yes	Yes
Provides freely password change facility	Yes	Yes
Provides U and GW-node mutual authentication	No	Yes
Provides GW-node and MS-node mutual authentication	No	Yes
Provides U and MS-node mutual authentication	No	Yes
Establishes secure session key between U and GW-node	No	Yes
Establishes secure session key between GW-node and MS-node	No	Yes
Establishes secure session key between U and MS-node	No	Yes

operation required at low powered MS-node. However, it is apparent from Table 5 that with extra memory capacity of 128 bits in smart card and some extra hash functions, the proposed scheme achieves higher performance. The most significant feature of our scheme is the establishment of mutual authentication and session key between every pair of the three participating entities.

8. Conclusion

A secure and efficient user authentication scheme is essential to offer reliable and proficient healthcare services *via* WMSNs. This work is motivated by the security problems of Kumar et al.'s scheme for healthcare services using WMSNs. In this paper, we have designed a user authentication scheme to eradicate the security problems of Kumar et al.'s scheme. Our scheme is user anonymous and is free from risks occurring due to loss of smart card of a user. It defies insider attack and password guessing attack. The most important feature of the scheme is that it establishes mutual authentication and provides session key between every pair of the participating entities, that is, user, GW-node, and MS-node.

Notations

U : User (professional)

U_a: Attacker

ID_u: Identity of U

PW_u: Password of U

SC : Smart card of U

GW-node: A powerful master node called Gateway-node

MS-node: Medical sensor-node

ID_g: Identity of GW-node

S_n: Identity of MS-node

J, K, Q: Long-term secret keys of GW-node

ID_{pt}: Identity of patient

M: Random nonce generated at the user side

T_g, T_g', T_{gs}, T_g'': Current timestamps generated at the user side

T_s, T_{sg}, T_s': Current timestamps generated at the GW-node side

T_a: Current timestamps generated at the MS-node-side

ΔT: Time interval for expected transmission delay

$h(\cdot)$: One-way hash function

S_y: Symmetric cryptographic operation (encryption/decryption)

$\oplus$: Bitwise Xor operator

$||$: Concatenation operator.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding of this research through the Research Group Project no. RGP-VPP-288.

References

[1] Y.-Y. Wang, J.-Y. Liu, F.-X. Xiao, and J. Dan, "A more efficient and secure dynamic ID-based remote user authentication scheme," *Computer Communications*, vol. 32, no. 4, pp. 583–585, 2009.

[2] M. K. Khan, S.-K. Kim, and K. Alghathbar, "Cryptanalysis and security enhancement of a 'more efficient & secure dynamic ID-based remote user authentication scheme'," *Computer Communications*, vol. 34, no. 3, pp. 305–309, 2011.

[3] M. Kumar, M. K. Gupta, and S. Kumari, "An improved efficient remote password authentication scheme with smart card over insecure network," *International Journal of Network Security*, vol. 13, no. 3, pp. 167–177, 2011.

[4] S. Kumari, M. K. Gupta, and M. Kumar, "Cryptanalysis and security enhancement of Chen et al.'s remote user authentication scheme using smart card," *Central European Journal of Computer Science*, vol. 2, no. 1, pp. 60–75, 2012.

[5] M. K. Khan, S. Kumari, and M. K. Gupta, "More efficient key-hash based fingerprint remote authentication scheme using mobile device," *Computing*, 2013.

[6] S. Kumari and M. K. Khan, "Cryptanalysis and improvement of 'a robust smart-card-based remote user password authentication scheme'," *International Journal of Communication Systems*, 2013.

[7] K. H. M. Wong, Z. Yuan, C. Jiannong, and W. Shengwei, "A dynamic user authentication scheme for wireless sensor networks," in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, vol. 1, pp. 318–327, Taichung, Taiwan, June 2006.

[8] H.-R. Tseng, R.-H. Jan, and W. Yang, "An improved dynamic user authentication scheme for wireless sensor networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 986–990, Washington, DC, USA, November 2007.

[9] M. L. Das, "Two-factor user authentication in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1086–1090, 2009.

[10] B. Vaidya, J. J. P. C. Rodrigues, and J. H. Park, "User authentication schemes with pseudonymity for ubiquitous sensor network in NGN," *International Journal of Communication Systems*, vol. 23, no. 9-10, pp. 1201–1222, 2010.

[11] M. K. Khan and K. Alghathbar, "Security analysis of 'two-factor user authentication in wireless sensor networks'," in *Proceedings of the 4th International Conference on Information Security and Assurance (ISA '10)*, vol. 6059 of *Lecture Notes in Computer Science*, pp. 55–60, Miyazaki, Japan, June 2010.

[12] M. K. Khan and K. Alghathbar, "Cryptanalysis and security improvements of 'two-factor user authentication in wireless sensor networks'," *Sensors*, vol. 10, no. 3, pp. 2450–2459, 2010.

[13] D. He, Y. Gao, S. Chan, C. Chen, and J. Bu, "An enhanced two-factor user authentication scheme in wireless sensor networks," *Ad-Hoc & Sensor Wireless Networks*, vol. 10, no. 4, pp. 1–11, 2010.

[14] S. Saleem, S. Ullah, and H. S. Yoo, "On the security issues in wireless body area networks," *Journal of Digital Content Technology and Its Applications*, vol. 3, no. 3, pp. 178–184, 2009.

[15] S. Ullah and K. S. Kwak, "Body area network for ubiquitous healthcare applications: *theory and implementation*," *Journal of Medical Systems*, vol. 35, no. 5, pp. 1243–1244, 2011.

[16] S. Ullah, H. Higgins, B. Braem et al., "A comprehensive survey of wireless body area networks—on PHY, MAC, and network layers solutions," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1065–1094, 2012.

[17] W.-Y. Chung, "Multi-modal sensing M2M healthcare service in WSN," *KSII Transactions on Internet and Information Systems*, vol. 6, no. 4, pp. 1090–1105, 2012.

[18] P. Kumar, S.-G. Lee, and H.-J. Lee, "E-SAP: efficient-strong authentication protocol for healthcare applications using wireless medical sensor networks," *Sensors*, vol. 12, no. 2, pp. 1625–1647, 2012.

[19] M. K. Khan, S. Kumari, and P. Singh, "Cryptanalysis of an 'efficient-strong authentication protocol (e-sap) for healthcare applications using wireless medical sensor networks'," *KSII Transactions on Internet and Information Systems*, vol. 7, no. 5, pp. 967–979, 2013.

[20] E. H. Callaway, *Wireless Sensor Networks, Architectures and Protocols*, Taylor & Francis, Boca Raton, Fla, USA, 2003.

[21] C. Chen, D. He, S. Chan, J. Bu, Y. Gao, and R. Fan, "Lightweight and provably secure user authentication with anonymity for the global mobility network," *International Journal of Communication Systems*, vol. 24, no. 3, pp. 347–362, 2011.

[22] Z. L. Ping and W. Yi, "An ID-based key agreement protocol for wireless sensor networks," in *Proceedings of the 1st International Conference on Information Science and Engineering (ICISE '09)*, pp. 2542–2545, Nanjing, China, December 2009.

[23] X. Lin, R. Lu, X. Shen, Y. Nemoto, and N. Kato, "SAGE: a strong privacy-preserving scheme against global eavesdropping for ehealth systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 365–378, 2009.

[24] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '99)*, pp. 388–397, Santa Barbara, Calif, USA, 1999.

[25] T. S. Messerges, E. A. Dabbish, and R. H. Sloan, "Examining smart-card security under the threat of power analysis attacks," *IEEE Transactions on Computers*, vol. 51, no. 5, pp. 541–552, 2002.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 916073, 14 pages
http://dx.doi.org/10.1155/2014/916073

Research Article

Wireless HDLC Protocol for Energy-Efficient Large-Scale Linear Wireless Sensor Networks

Daniel Mihai Toma, Joaquin del Rio, and Antoni Mànuel

SARTI Research Group, Electronics Department, Universitat Politècnica de Catalunya (UPC), Rambla Exposició 24, Vilanova i la Geltrú, 08800 Barcelona, Spain

Correspondence should be addressed to Daniel Mihai Toma; daniel.mihai.toma@upc.edu

Received 31 October 2013; Revised 10 February 2014; Accepted 16 February 2014; Published 17 April 2014

Academic Editor: Zuqing Zhu

Wireless sensor networks (WSNs) have been widely recognized as a promising technology that can enhance various aspects of infrastructure monitoring. Typical applications, such as sensors embedded in the outer surface of a pipeline or mounted along the supporting structure of a bridge, feature a large-scale linear sensor arrangement. In this paper, we propose a new bidirectional wireless communication scheme, based on the high-level data link control (HDLC) standard, for devices with short-range transmission capabilities for linear sensor topology. By applying for the first time a standard data layer along with a time division multiple access (TDMA)-based medium access control (MAC) and time synchronization technique specifically designed for the linear topology, we address the interoperability problem with guaranteed energy efficiency and data link performance in linear sensor topology. The proposed Wireless HDLC supports half-duplex communication, point to point (peer to peer), and multipoint networking.

1. Introduction

The existing applications of WSNs span a very wide range, including remote system monitoring and control, fraud detection, environmental monitoring, and so forth. So far, little focus has been given to low-power WSNs for linear topologies. Some papers present linear wireless networks for bridge [1], pipeline [2], or overhead transmission lines [3–5] application. For instance, real time monitoring of power cables thermal conditions and ambient conditions such as temperature and humidity could result in higher dynamic rating of transmission lines and will increase the utilization of their power-carrying capabilities [6]. However, the systems described in these works are proprietary solutions, following no particular standard for communication and arising interoperability problems when they are used. Due to their linear geometry, direct transmission from data source to sinks is usually not practical because the sensor nodes (SNs) have a limited communication range and data sources are generally far away from the sinks. Therefore, a multihop network is a good choice for data routing, and clustering topology is appropriate to achieve network scalability [7]. Topology design, power usage minimization, and installation cost are very important for successful deployment of linear WSNs while meeting the application requirements. This paper proposes a new WSN technology based on standard HDLC protocol for long-term continuous monitoring of large-scale linear infrastructures so that efficient monitoring and management systems can be established.

As shown in Figure 1, wireless multifunctional SNs are installed on the critical components of large-scale infrastructures such as bridges, pipelines, or power cables, using linear topology. However, the power supply constraints of the WSNs deployed in these infrastructures pose great challenges in energy consumption. Hence, there is a need for reliable and low-power WSNs for linear networks capable of being powered through harvesting devices for long-term monitoring. Therefore, WSNs based on IEEE 802.15.4 standard for low-power wireless transceiver technology need to be used [8]. Generally, the transmission range of the nodes is assumed to be 10–100 m with data rates of 20 to 250 kbps [9]. Hence, large network and multihop communication are required

Table 1: Comparison in terms of devices power consumption, hops limit, and topologies for existing standard protocols and the proposed Wireless HDLC.

Protocol	ZigBee PRO	WirelessHART	6LoWPAN	SP100.11a	Wireless HDLC
Power consumption	Very low in end devices High in router devices	Low in all WirelessHART field devices	Low in host devices. High in router devices	Low in end devices. High in router devices	Low in all field devices
Maximum number of hops	30	4	255 (theoretical)	255 (theoretical)	124 normal (8 bit address) 8190 in extended (16 bit address)
Main topologies	star, mesh	mesh	mesh	mesh	bus, star

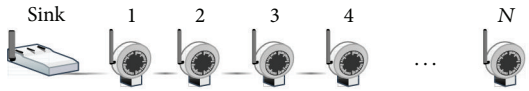


Figure 1: Wireless linear network as a chain of low-power sensor nodes (SNs) deployed in linear topology.

so that nodes relay the information to the data collector, that is, the sink. Moreover, these networks have to combine power and routing awareness, communicate power efficiently through the wireless medium, integrate data with networking protocols, and promote cooperative efforts of SNs [10].

Several wireless standards such as ZigBee [11], WirelessHART [12], 6LoWPAN [13], and SP100.11a [14] formed on top of IEEE 802.15.4 standard, which specifically address the typical needs of wireless control and monitoring applications have been actively pushing the application of wireless technologies in industrial measurement and control applications. The ZigBee protocol is the driver for the development of the 802.15.4 standard and uses the IEEE 802.15.4 PHY and MAC layers but it defines the network and application layer. The WirelessHART and ISA100.11a also use the 802.15.4 PHY but define their own MAC and network and application layers [15]. Following the OSI reference model, it specifies the network and transport layer and use the 802.15.4 PHY and MAC sublayer of the data link layer [16]. Generally, the network topology of these standards is designed as a mesh network and enables application-specific solutions to be developed for WSNs. However, except for the WirelessHART network where each device can act as a source or a router, in the other standard networks, dedicated routers nodes are necessary to provide communication between source nodes and sink. Moreover, the number of network hops allowable within these standards is limited. The maximum distance from a node to the sink that is allowed by the WirelessHART is of 4 hops, by ZigBee and ZigBee PRO is of 10 hops and 30 hops, respectively, and ISA permit up to 20 hops. The theoretical number of hops limit in a 6LoWPAN network is 255 (8 bits hop limit field). However, 6LoWPAN networks with this specification have not been reported in the literature to date [16].

These limitations are very important for WSNs deployed in large-scale infrastructures, where the number of hops is assumed to be in order of hundreds and each node is both source and router. In this framework, we present the implementation and evaluation of a bidirectional wireless communication schema for linear IEEE 802.15.4-compliant WSNs based on HDLC standard [17] (Wireless HDLC). Table 1 illustrates the capabilities of the proposed Wireless HDLC network in comparison to the existing standard WSN solutions.

The Wireless HDLC adopts the IEEE 802.15.4 PHY layer but defines a new TDMA-based MAC, network and transport layers based on HDLC standard. The issue regarding synchronization of nodes throughout the network is addressed by applying any of the time synchronization techniques available such as TPSN (timing-sync protocol for sensor networks) [18] or PTP (precision time protocol) [19]. These techniques may exchange timestamp messages to synchronize distributed clocks in a network while meeting the power usage and bandwidth minimization required by WSNs. The chain of short-ranged wireless sensors creates a virtual wired link by means of an ad-hoc network. The system does not require complex routing techniques. The proposed Wireless HDLC supports half-duplex communication providing a bidirectional link between the SNs and the sink. The communication is done in rounds, one time from the sink node to the last node in the network (end node) and one time from the end node to the sink. The bidirectional link acting as a virtual conveyor belt can be used to collect data from different sensors along thepath or send data from the base station to different sensors in the network. The data from multiple devices is encoded as HDLC frames and is collected in the available space of the IEEE 802.15.4 standard packet up to a maximum size of 125 octets [8]. In this way various nodes can send variable length packages in one communication round of the transmission grid, following a standard form.

The paper is organized as follows. In Section 2 we review some existing related works. Section 3 describes the Wireless HDLC protocol stack layered architecture based on HDLC standard. Section 4 validates our design by demonstrating the Wireless HDLC network. Finally, this paper is concluded and the future work is presented in Section 5.

2. Related Work

Wireless sensor network has been the focus of extensive study recently [20–24] proposing a wide variety of algorithmic and communication protocols solutions. Most studies are focused

on generic WSNs which assume that sensors are deployed randomly and abundantly in the same area and perform the same function. The smart features envision in the roadmap for research and development of the next generation WSNs, as digitalization, flexibility, intelligence, and customization deal with the interoperability problem in order to allow plug-and-play capability to accommodate progressive technology upgrades with hardware and software components. The plug-and-play capability for low-power WSNs can only be achieved by applying standard technologies such as ZigBee, WirelessHART, 6LoWPAN, or ISA SP100. Although, these standard solutions give support for mesh networking and can be used in a wide range of applications such as home automation, smart energy, building automation, industrial automation, and personal health care, they are not optimized to work in large-scale areas. Due to the geometry of large-scale infrastructures, WSNs with linear topology have to be used.

However, the design of these WSNs pose several challenges due to their linear nature, limited energy source, robustness to dynamic environment, and scalability to numerous number of sensor nodes. What follows describes the WSN design challenges for an efficient communication in this environment.

2.1. Energy Consumption. A wireless module is equipped with a limited energy source (i.e., battery, harvesting and device) and hence has an energy consumption capacity and lifetime that is dependent on that source. In a WSN, each node plays two separate and complementary roles: it can originate data and also has to route data. Moreover, if a few nodes deplete their energy resources, it can cause significant topological changes and might require rerouting of packets and reorganization of the network.

2.2. Operating Environment and Fault Tolerance. WSNs for large-scale infrastructures have to be designed with extreme environments in mind. The environmental interference but also physical damage or a depleted energy source may cause an SN to fail. However, the failure of a single node should not affect the overall operation of the network.

2.3. Scalability and Network Topology. Depending on the infrastructure's length and the number of points of interest, the quantity of SNs deployed in these WSNs may be in the order of hundreds. Therefore, WSN protocols have to be designed to work with these large numbers of nodes. Also a major challenge is the deployment of these SNs to minimize the cost of deployment under the constraint of coverage, connectivity, and link outage probability so that the phenomenon of interest can be monitored efficiently. Moreover, additional SNs can be redeployed at any time to replace the malfunctioning nodes or due to changes in task dynamics.

Between all these difficulties, the topology is maybe the most important challenge facing the development of WSNs for large-scale infrastructures. Topology is important for any type of network because it has a great impact on the communication performance of the system. In the literature we can find some examples of algorithms and protocols that are specifically aimed for linear topologies. In [25] Zimmerling et al. proposed the Minimum Energy Relay Routing (MERR) algorithm. Its aim is to minimize the routing path from every node to a common control center. In particular that work covers the routing problems of the special case of a linear network where nodes are located close to their neighbors. The Directional Scheduled MAC (DiS-MAC) [26] is another protocol that has been developed for WSNs that shows a linear topology. With DiS-MAC, Karveli et al. a fail tolerant unidirectional routing protocol for linear network is proposed. In [23] the Wireless Wire (WiWi) protocol is described. This work proposes a bidirectional wireless communication schema with deterministic properties in terms of throughput and latency over a strip of pervasive devices with short-range transmission capabilities. The system is synchronous and fault tolerant and can provide support for an end-to-end communication.

As shown in Table 2, both MERR and load balanced routing are focused on routing problems without considering the underlying MAC protocol or the transport layer. The DiS-MAC protocol is focused on TDMA-based MAC protocol and also on routing problems in the linear network but suffers from being unidirectional. The WiWi protocol defines a TDMA-based MAC protocol and it provides bidirectional communication over a single RF channel. However, in this protocol, power consumption is not considered, and the SNs have to be in active mode for long periods of time. Moreover, compared to standard technologies such as ZigBee, WirelessHART, 6LoWPAN, or ISA SP100, these protocols have a significant drawback providing incomplete architecture and no standard is followed. In next generation WSN's vision, large, integrated, complex systems require different layers of interoperability, from a plug or wireless connection to compatible processes and procedures [27].

Although it bears many similarities with these works, such as routing techniques, the implementation of Wireless HDLC network differentiates itself from them in many other aspects. In particular, different from the protocols proposed in these papers is that HDLC has been widely implemented in cable networks because it supports both half-duplex and full duplex communication lines, point to point (peer to peer) and multipoint networks and switched or nonswitched channels. The HDLC protocol resides with layer 2 of the OSI 7-layer communication model, the data link layer. However, it offers three different modes of operation supporting a reliable and orderly transfer of packets in a distribute network (transport layer) and specifies the types of stations for data link control. The HDLC is essentially a centralized wireless network which uses a central network manager, the primary station, to provide routing and communication schedules to meet the requirements of wireless applications.

The design of Wireless HDLC architecture is based on the requirements of WSNs for large-scale infrastructures, meeting the challenges for an efficient communication in these environments and the need for standard solution for next generation WSNs applications. In the proposed Wireless HDLC network, frames of many SNs can be transmitted in

Table 2: Several protocols for linear WSNs.

Protocol	Communication	OSI Layer
MERR [25]	Unidirectional	Network layer
DiS-MAC [26]	Unidirectional	TDMA-based MAC and network layer
WiWi [23]	Bidirectional	TDMA-based MAC and simple network layer
Load balanced routing [35]	Unidirectional	Network layer

one IEEE 802.15.4 standard packet traveling in the linear networks from the sink node to the end node and from the end node to sink node. This allows data collection and control of individual or groups of low-power field devices deployed with linear topology.

3. Detailed Description

This section presents a unique vision for the WSNs with linear topology in which the wireless network provides half-duplex communication and point-to-multipoint connection. Also, following the OSI reference model, the protocol stack layers of the Wireless HDLC networks are described. In Figure 1 the proposed Wireless HDLC network with a central network manager (sink) and SNs with routing functionality capable of building up the linear link is illustrated.

3.1. Data Link Layer. The Wireless HDLC protocol stack is designed as a synchronous multihop communication scheme and uses a TDMA-based MAC protocol that provides collision-free multiple access. The decision to focus on the TDMA approach arises largely from the fact that the energy consumption is significantly lower than the contention-based due to collision-free communication and minimization of idle listening. The operation of TDMA-based MAC is divided into sequences of phases as depicted in Figure 2.

The sequences of phases consist of a network set-up phase and a communication phase. The set-up phase is designated for updating the sensor model at the sink. The sink informs each node about slots in which it should listen to other nodes' transmission and about the slots, which the node can use for its own transmission. Once the linear network is built, the system enters into the data transfer phase. The data transfer phase is divided into up-session and down-session. Each session consists of a data transmission/reception period and an idle period. The up-session provides time slots for each node, starting with the sink up to the last node in the network. The down-session provides time slots for each node, starting with the last node to the sink. Assuming that there are N nodes within a network then the session's period consists of exactly N slots. The slot duration is the time required to transmit two times a maximum sized IEEE 802.15.4 packet (10 ms) and the duration of idle period depends on the network communication rate. Any node that does not have data to send is assumed to be a repeater. During the data transmission/reception period, each node turns on its radio and sends/receives its data to/from his neighbor over its allocated slot-time and keeps its radio off at all other times.

Time synchronization is a fundamental issue in TDMA protocol and for this, accurate timing is ensured by applying synchronization techniques such as PTP protocol [19]. By exchanging timing messages, within HDLC frames, the network may achieve high synchronization precision, in order of microseconds or submicroseconds, as an exact measurement of the transmission and reception times.

The data traveling in the Wireless HDLC network is a collection of HDLC frames (superframe) transmitted in an IEEE 802.15.4 packet as illustrated in Figure 3. Each HDLC data is organized into frames using a frame delimiter, or flag, which is a unique sequence of bits [01111110]. A frame delimiter at the end of an HDLC frame may also mark the start of the next HDLC frame. Each frame contains an address field, a control field, information data, and optionally, a CRC field. The length of the address field depends on the total number of wireless sensors in the network and is normally 8 or 16 bits in length for networks of maximum 124 nodes, respectively, 8190 nodes. As specified in the ISO/IEC-13239 standard [17], the control field is 8 bit length indicating a command or a response and a numbers sequence where applicable. The standard HDLC frames indicated through the control field are the information transfer format command and response (I format), used to transmit user data between stations; the supervisory format commands and responses (S format), used to perform control functions; and the unnumbered format commands and responses (U format), used for control purposes. The 5th bit position in the control field is called the poll/final bit or p/f bit and is used to provide dialogue between the primary station and secondary stations. This is a very useful control in linear WSNs, where the primary station can use the poll function to acquire responses from the SNs (secondary stations) in an end-to-end transmission.

The data transfer transactions in the Wireless HDLC network are designed based on HDLC specifications. The HDLC offers three different modes of operation, the Normal Response Mode (NRM), Asynchronous Response Mode (ARM), and Asynchronous Balanced Mode (ABM) [17]. However, because of the linear nature and the requirements of Wireless HDLC only NRM and ABM modes are used.

The NRM is a master-slave mode and is used to coordinate the data transfer between the network nodes and the sink. In NRM, the sink (primary station) gives permission for each network node (secondary station) to speak. The secondary station can only transmit a response when, and only when, it is instructed to do so by the primary station. This transmission from the secondary station to the primary station may be one or more information frame. Once the last frame is transmitted by the secondary station, it must wait once again for explicit permission to transfer anything from the primary station.

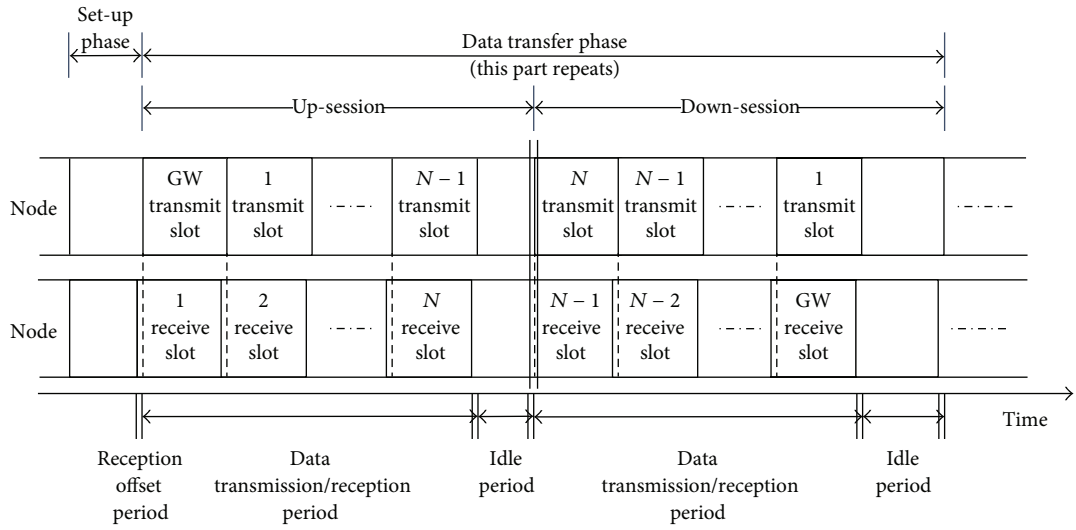


Figure 2: TDMA-based MAC protocol for Wireless HDLC networks.

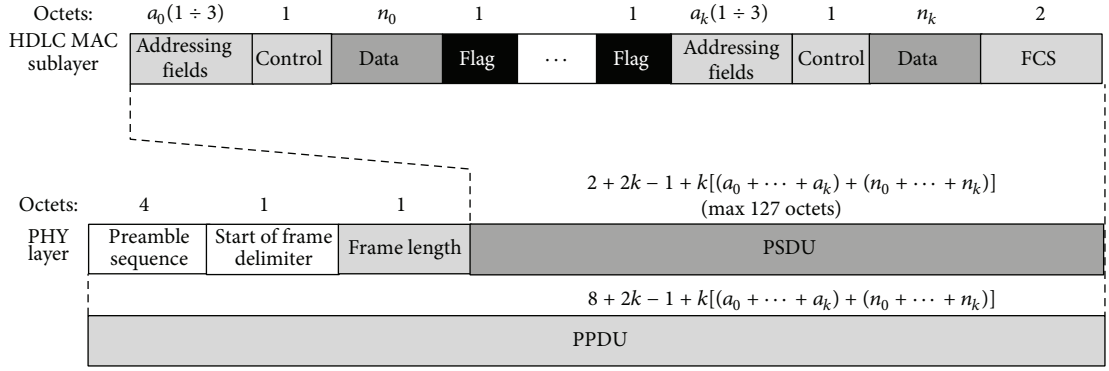


Figure 3: Schematic view of HDLC superframe.

In Figure 4, an example of Normal Response Mode (NRM) data transfer is described for the wireless multipoint line using the P(oll)/F(inal) procedure. In this example, a primary station (sink) and 5 secondary stations (SNs) on a multipoint line are illustrated. The sink sends out its data in the TDMA up-session (1) with $P = 0$ to SN 5 (C5) and with $P = 1$ to SN 2 (C2); C2 can transmit data. Next, in down-session (1), C2 sends data to primary station with $F = 0$ to a signal that has more data to send in the next down-sessions. Moreover, the sink can receive data from various secondary stations in the same session. Therefore, in the up-session (2) the sink informs C1 that can transmit data. Hence, in the down-session (2), the sink receives data from C1 and C2. However, the total length of these frames cannot exceed the maximum length of an IEEE 802.15.4 standard packet. In the wireless HDLC architecture the primary station is also the source of time synchronization. Each IEEE 802.15.4 packet transmitted by the primary station contains a frame with its timing information. Using this data, the secondary stations can adjust their clock to the primary station.

The ABM mode allows the data transfer between two peer devices (combined stations). The combined station acts as both a primary and a secondary station. The ABM is used as data transfer transaction mode between neighbor nodes in the set-up phase and also in the communication phase, when a node transmits rescheduling or energy information to its neighbors.

3.2. Network Layer. In the literature, many routing algorithms and protocols are reported for WSNs. As reported by Singh et al. all major routing protocols proposed for WSNs are broken down into seven categories: location-based protocols, data-centric protocols, hierarchical protocols, mobility-based protocols, multipath-based protocols, heterogeneity-based protocols, and Quality of Service (QoS)-based protocols [28].

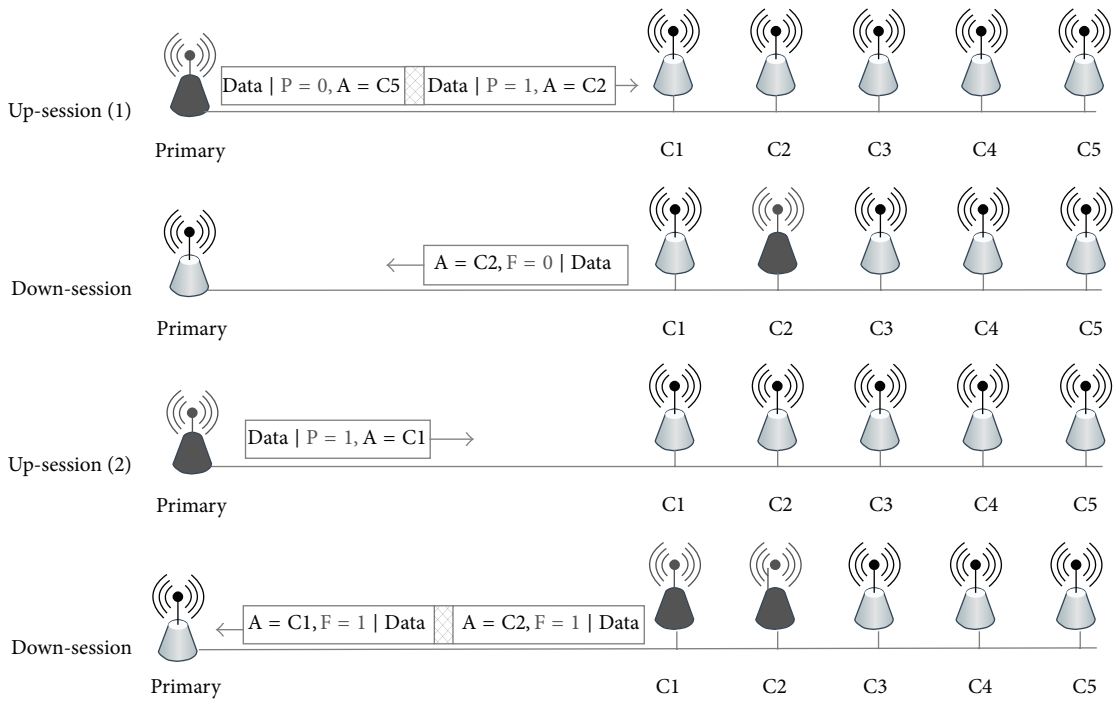


Figure 4: Data transfer in Wireless HDLC multipoint line using poll/final procedure.

From the literature review of the routing techniques and based on the network topology and challenges of the Wireless HDLC, the hierarchical protocol Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [29] was chosen. The main reason for adopting this routing protocol was that in Wireless HDLC, where wireless network is constituted by low-power SNs distributed along a strip, the PEGASIS can form a chain from SNs so that each node transmits and receives data from a neighbor. PEGASIS protocol provides the mechanism for sensors to transmit data to their closest neighbor until the end of the network is reached. In the PEGASIS routing protocol, the construction phase assumes that all the sensors have knowledge about their positions and use a greedy approach. In the Wireless HDLC, the positions of the sensors can be determined through the rules such as best link quality or lowest delay. When a sensor fails or dies due to low battery power or when it has run out of energy, the chain is constructed using the same greedy approach by bypassing the failed sensor.

3.3. Transport Layer. The transport control protocols for WSNs are important for reliable data dissemination and energy-conservation. Generally, transport control protocols may include two main functions: congestion control and loss recovery [18]. In order to weaken congestion, the transport layer can use end-to-end mechanism like TCP or hop-by-hop approaches. However, the end-to-end mechanism, which is based on acknowledgments and end-to-end retransmissions, imposes significant overhead for the implementation of these solutions in WSNs. Moreover, hop-by-hop approaches can control congestion with less ongoing packets in networks, while it needs to change the behavior of each node on the way from source to destination. Due to bad quality of wireless channel, sensor failure, and/or congestion, the transport layer should manage the packet loss and data from SNs should be reliably transferred to the sink. Also, the commands and queries from sink should be reliably delivered to the target SNs to assure the proper functioning of the WSNs.

This transport layer aims to address both the reliability and congestion problems using a mixed solution between end-to-end mechanism and hop-by-hop approach through HDLC supervisory frames. Between a pair of nodes, which transmit data accordingly to PEGASIS routing technique, in case of packet errors, each node performs hop-by-hop recovery to fetch the lost packets from neighbor nodes. An SN tries to transmit a package to its neighbor until it receives a Receive Ready (RR) supervisory frame, which acknowledges the reception of this package. Using the TDMA-based protocols explained above, the number of retries is limited by the time slot period and the size of the superframe. To obtain a good compromise between the hop-by-hop reliability, the link delay, and the energy consumption in the proposed WSN, the transmission time slot was set to 10 ms permitting two reties of a maximum sized IEEE 802.15.4 packet.

However, end-to-end reliability cannot be maintained in all cases even though hop-by-hop reliability is ensured. Therefore, an end-to-end mechanism is also used for the data transfer between SNs and the sink as illustrated in Figure 5. This method is based on the Selective Reject Automatic Repeat Request (SREJ ARQ) technique provided by HDLC protocol. As shown in Figure 5, the SREJ ARQ technique uses the Send Sequence Number (NS) and the Receive Sequence Number (NR) to control the correct data exchange between the sink and a Sensor Node (SN).

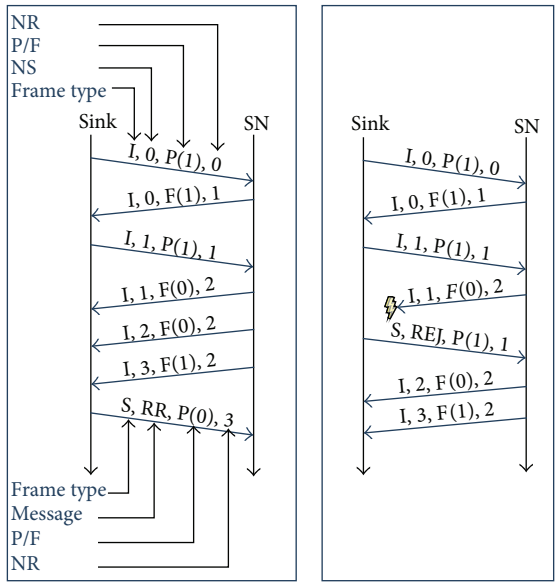


Figure 5: HDLC operation: (a) normal two-way data exchange and (b) selective reject ARQ.

3.4. Application Layer. Generally, the application layer provides necessary interfaces to the user to interact with the physical world through the WSN. Hence, the role of the application layer is to abstract the physical sensors and topology of the WSN for the applications. The large variety of available sensor types complicates the integration of sensors into monitoring systems. Therefore, the use of standard interfaces and data encodings such as IEEE 1451 Smart Transducer Interface Standard [30, 31] and time synchronization standards such as IEEE 1588 PTP protocol [17] is recommended. Hence, we propose the use of standard applications, which can be easily implemented on the Wireless HDLC network, in order to hide the underlying layers, the network communication details, and heterogeneous sensor hardware from the applications built on top of it.

Figure 6 illustrates the proposed half-duplex communication model suitable for low-power WSN and the handling of downstream and upstream data flows. The upstream data flow is generated by the sink and flows up to the last node in the network. The downstream data flow is generated by the last node in the network and starts when the upstream flow reaches this node. In this way a two-way communication cycle is generated which repeats itself along the time based on a network communication period. This data flow has a strict staggered pattern: each node will calculate the trigger time for receiving and transmitting a time slot based on the total number of hops. The hop-by-hop approach in a pair of consecutive nodes is also depicted in Figure 6; a node transmits an IEEE 802.15.4 packet to the next node and waits for a message to acknowledge this. With a transmission rate of the 802.15.4 Standard 2.4 GHz physical layer of 250 kb/s, a message with the maximum size of 127 bytes (125 bytes payload + 2 bytes PHY dependent for IEEE 802.15.4) transmitted between pair nodes in 4000 us and with an acknowledge timeout of 1000 us is allowed for one retransmission. Depending on the synchronization accuracy, the pairs of sensors will use a time offset for the transmit-receive process to avoid loss of messages.

The upstream IEEE 802.15.4 packet transmitted by the sink node contains HDLC frames with data, commands, or timing information to SNs. The downstream packet is a collection of HDLC frames with the responses from SNs to the sink commands. If an SN does not receive any command from the sink node it will work as a router and will transmit the downstream packet to the next hop.

4. Implementation and Performance Evaluation

This section starts by introducing the hardware platform we use to evaluate the Wireless HDLC and then offers a description of its implementation and provides the results of our proposal.

4.1. Hardware Platform. We base our implementation on the XBee and XBee-PRO 802.15.4 OEM RF modules [32] provided by Digi. The modules contain the MC13211 platform which incorporates a low-power 2.4 GHz radio frequency transceiver and an 8 bit Freescale HCS08 microcontroller [33]. The modules are driven by a 16 MHz crystal source and contain an internal event timer block of 24 bit clocked at a rate varying from 15 kHz to 2 MHz used to maintain the SNs time and synchronization.

Freescale provides for this platform a simple IEEE 802.15.4 physical layer library in ANSI C used to build our proposed protocol stack and the new firmware for XBee modules. A prototype design, using the XBee module, has been developed for the Wireless HDLC. The prototype is

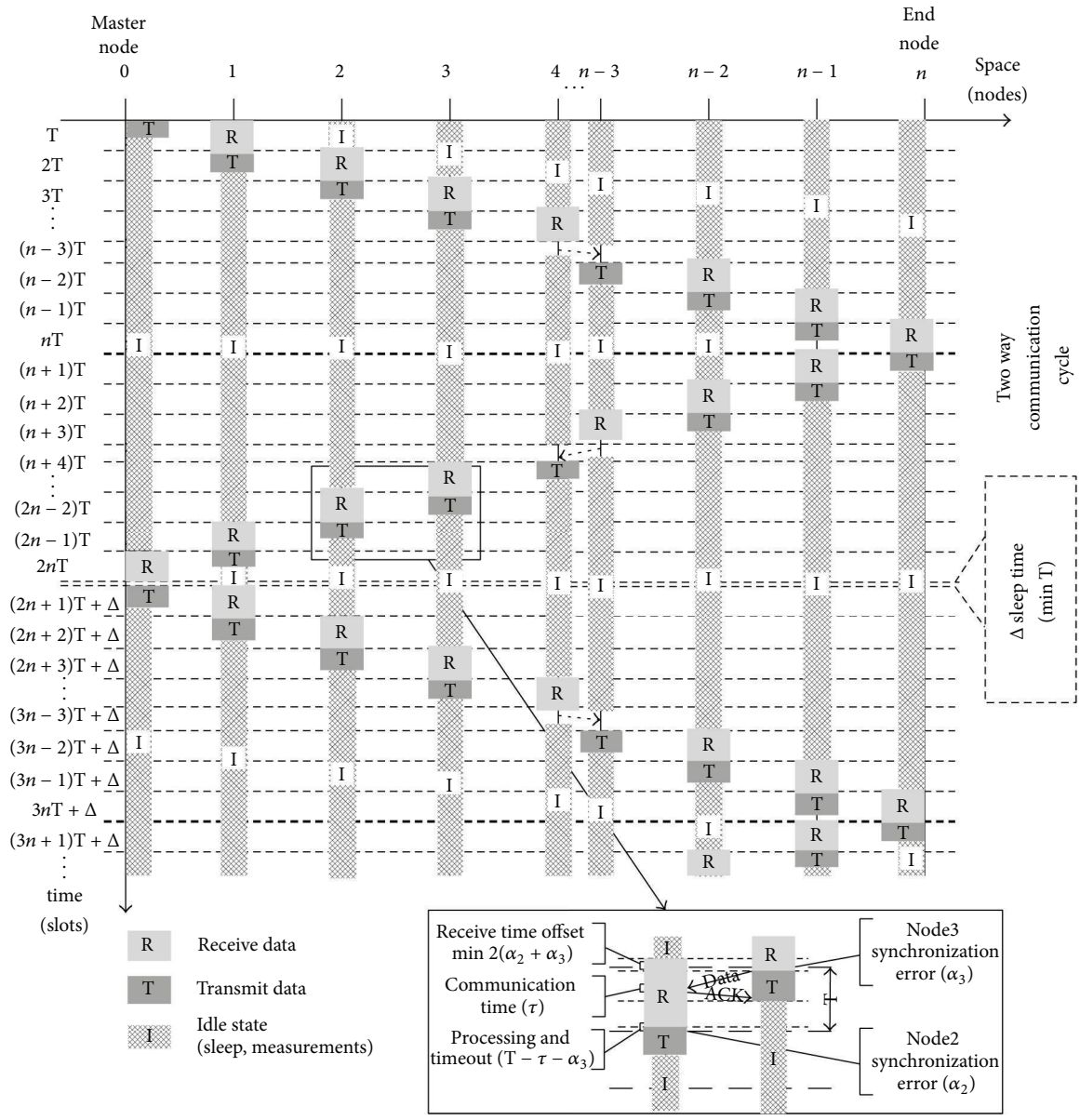


Figure 6: Bidirectional communication for low-power Wireless HDLC networks.

equipped with a harvesting device as power supply [34]. Also, the board includes 8 ADC channels of 10 bit, one SPI port, and a RS232 interface. Hence this platform can be equipped with various sensors to monitor phenomenon of interest.

4.2. Protocol Stack Design. Figure 7 describes the overall design of the Wireless HDLC data link layer which consists of four major modules: timer, TDMA function, HDLC framer, and the state machine. The timer module provides accurate timing to ensure the system operates correctly. The timer module has been designed as a real time clock capable of providing accurate triggers for wake-up, measure, receive, or transmit states of the SNs and also to keep the 10 ms time slots in synchronization. To keep the synchronization of the timer module, a simplified PTP protocol [34] is used to exchange timing messages in the network. This protocol provides methods to implement transparent clocks in multihop networks, to calculate the delay time, and to adjust the offset between the master clock and slave clocks. With this method, each SN can also synchronize their source clock using the trim capability of the 16 MHz crystal, which reduces their drift with the master source clock. Using the 24 bit counter clocked at a rate of 2 MHz, the nodes are able to obtain synchronization accuracy in the order of a microsecond.

The TDMA function determines the usage of time slots for communication. This module uses a schedule to calculate the number of time slots between the downstream and upstream data flows for each sensor according to the number of nodes in the network. It also calculates the time for the next communication cycle based on network communication

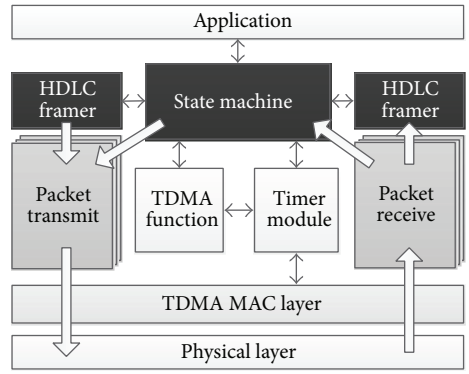


Figure 7: HDLC protocol stack architecture.

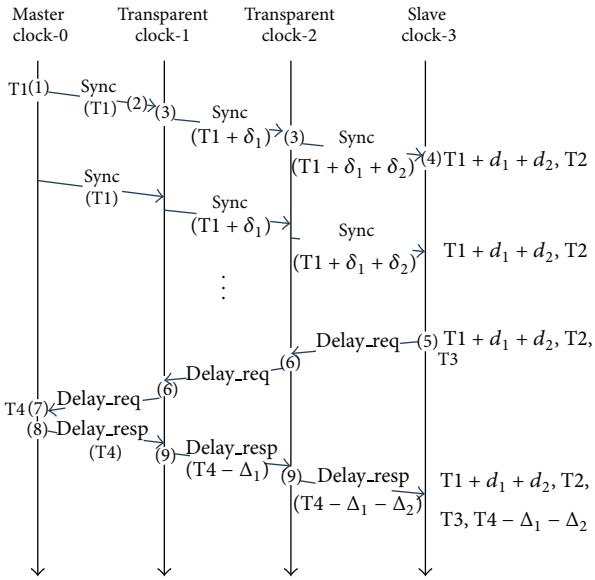


Figure 8: Synchronization messages exchange for linear topology. The slave communicates with the master through transparent clocks.

period. Every event that can affect the time scheduling will cause the TDMA function to reassess this scheduler.

The functionality of the HDLC framer is to decode the received messages in order to get the relevant information for the SN and to generate and add HDLC frames to the transmitted messages. If an SN receives an HDLC frame from the sink node with his unicast address, it will delete the frame from the IEEE 802.15.4 packet and will retransmit the remaining package. Next, in the downstream data flow, the node will add the response, for the sink query, to the received superframe and it will transmit the new package to the next node in the network.

The state machine is responsible for sending and receiving a packet over the transceiver. Based on the TDMA triggers, it executes the transaction in a slot and uses a light implementation of PTP protocol to adjust the timer module.

4.3. PTP Protocol over Wireless HDLC. The PTP standard specifies a clock synchronization protocol applicable to distributed systems consisting of one or more nodes communicating over a network. The protocol provides a mechanism for synchronizing the clocks of participating nodes to a high degree of accuracy and precision. However, a full implementation of this protocol as is specified by the IEEE 1588 standard is not suitable for low-power WSNs. The main challenge is that the IEEE 1588 messages size is more than the maximum size allowed in IEEE 802.15.4. Therefore, a light implementation of this standard is proposed for Wireless HDLC networks. The wireless implementation of PTP protocol is built on top of HDLC protocol consisting of PTP devices including ordinary clocks and transparent clocks. Figure 8 illustrates the basic pattern of synchronization message exchange in the linear topology of Wireless HDLC.

The timestamp information obtained with this exchange of PTP messages may be used to compute the offset of the SNs (transparent clocks) with respect to the sink (master clock) and the mean propagation time of messages between the two clocks. Considering that the master-to-slave delay is equal to the slave-to-master delay and by using the time information in the four messages illustrated in Figure 8, these quantities can be calculated from the measurement values $T1$, $T2$, $T3$ and $T4$:

$$\begin{aligned}\text{delay}_k &= \frac{\left(T2 - T1 + \left(\delta_1 + \cdots + \delta_{k-1}\right)\right)}{2} \\ &\quad + \frac{\left(T4 - \left(\Delta_1 - \cdots - \Delta_{k-1}\right) - T3\right)}{2}, \\ \text{offset}_k &= \frac{\left(T2 - T1 + \left(\delta_1 + \cdots + \delta_{k-1}\right)\right)}{2} \\ &\quad - \frac{\left(T4 - \left(\Delta_1 - \cdots - \Delta_{k-1}\right) - T3\right)}{2},\end{aligned} \tag{1}$$

where $\delta_1, \ldots, \delta_{k-1}$ is the accumulated residence time in transparent clocks for Sync message and $\Delta_1, \ldots, \Delta_{k-1}$ is accumulated residence time in transparent clocks for Delay_Req message. The synchronization interval is equal to the network

Table 3: PTP message fields in HDLC frame.

Bits								Octets
7	6	5	4	3	2	1	0	
			HDLC flag(01111110)					1
			HDLC address					1 ÷ 2
			HDLC control (0 | 0 | 0 | P/F | 1 | 0 | 1 | 1 for PTP)					1
Two step flag			PTP control					1
			Origin timestamp (optional)					8
			HDLC CRC (optional)					2
			HDLC flag(01111110)					1

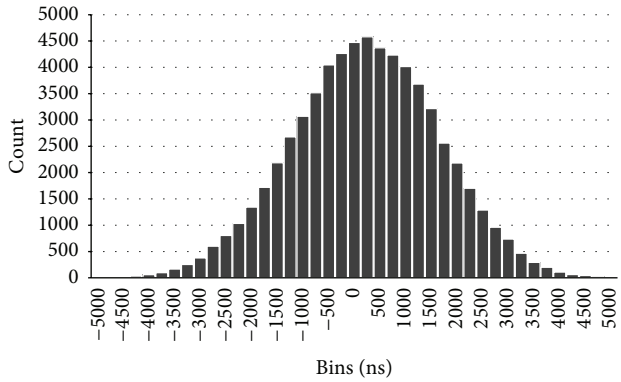


Figure 9: PPS signals delay histogram in linear network between the four consecutive SNs.

sampling rate. A Sync message is sent by sink node in each upsession. The delay request interval can be adjusted to a lower frequency to reduce energy consumption and the overhead (i.e., a Delay_Req message is sent by the last node in the linear network for every ten Sync messages).

The general format of the PTP messages codified as an HDLC frame is shown in Table 3 and has a size of 14 bytes (8 bit address) or 15 bytes (16 bit address).

Each PTP message starts and ends with an HDLC flag. The PTP messages are defined as unnumbered frames (U-frames) using a nonstandardized control value of the HDLC control field (0 | 0 | 0 | P/F | 1 | 0 | 1 | 1). The next field in the frame indicates through "Two step flag" if the Follow_Up message is used and what type of PTP message is transmitted: Sync, Follow_up, Delay_Req or Delay_Resp.

The histogram presented in Figure 9 represents the time deviations between the sink node and the first four consecutive SNs in Wireless HDLC. The mean time delay is 1144.30 ns with a standard deviation of 1082.27 ns and a maximum value of 4506 ns.

The obtained results show that through this technique we are able to obtain microsecond precision with efficient energy saving. Hence the SNs can maintain their TDMA-based protocol with very high accuracy, providing collision-free communication and minimization of idle listening.

4.4. Network Demonstration. Based on the prototype stack, we develop a linear self-powered WSN, with the aim of building a demonstration network for monitoring the environmental and infrastructure parameters. The demonstration network contains one sink (master) node and seventeen SNs. The SNs have been designed as alert nodes to detect critical events and generic monitoring nodes to measure different parameters such as infrastructure components temperature, ambient temperature, and humidity. The SNs in the network have been deployed with a gap varying from 50 m to 100 m that can cover approximately 1.2 km of infrastructure as shown in Figure 10. Each node can communicate with two nodes to the left and two nodes to the right. If, for example, node 10 is damaged, node 11 can still send the downstream data flow through node 9. The SNs have been programed to respond to different types of commands like "take sample," "get calibration," or "get capability." To reduce the amount of information transmitted over the air, these sets of commands have been encoded in an 8 bit length format. If a user or the master node transmits the "take sample" command in the upstream data flow to different nodes using unicast addresses or multicast addresses, these sensors will put their raw measurements in the downstream data flow.

Since the Wireless HDLC works on top of IEEE 802.15.4 packets (125 bytes), the master node can only read the measurements from the entire network in two communication cycles. In this case we chose to use one cycle to collect data from the first 7 line devices and the alert sensors and another cycle to collect data for the last line devices and the alert sensors. This pattern is normally used to perform near real time measurements using a communication period of 5 s. In this way the self-power SNs can stay in sleep mode between cycles of communication to reduce the power consumption.

4.5. Qualitative Evaluation. The Wireless HDLC network is a synchronous network and can provide deterministic and predictable latency and throughput in both directions. This synchronization also provides additional capabilities for energy management by turning off the SNs during idle time. Because the times when a node should sense and communicate are clearly identified, the SNs can be completely turned off without affecting the connectivity of the network.

Table 4: Power consumption measurements (Xbee PRO characteristics: TX-215mA, RX-55mA, and Active-7.5mA).

	Energy consumption (mJ)
Sink-to-SN send (a)	0.831 + 0.046 × size
SNs-to-sink receive (b)	0.641 + 0.046 × size
Multicast send (c)	0.625 + 0.079 × multicast + 0.046 × size (multicast < nodes)
Multicast receive (d)	0.435 + 0.079 × nodes + 0.046 × size
Point-to-point send and receive (WirelessHART) (e)	0.974 + 0.046 × size

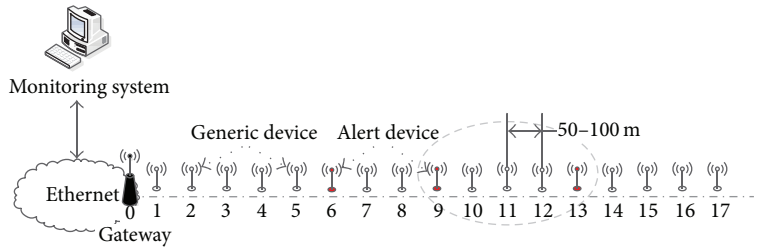


Figure 10: Each node can communicate with four nodes.

Based on formula reported by Toma et al. the communication energy consumption for a wireless node can be modeled as follows [36]:

$$E_C = N_T \left[P_{LO}\left(t_{tx} + t_{st}\right) + P_{PA}\left(t_{tx}\right)\right] + N_R \left[P_R\left(t_{rx} + t_{st}\right)\right], \quad (2)$$

where P_{LO} is the transmitter power consumption, t_{tx} is the transmitter on time, t_{st} is the transmitter start-up time, P_{PA} is the transmitter output power, P_R is the power consumed by the receiver, t_{rx} is the receiver on time, t_{st} is the receiver start-up time, and N_T and N_R are the number of times the transmitter and receiver are switched on per unit time, respectively. However, as shown in Figure 11, two other energy consumption components are added to model (2) above. The final communication energy consumption model for the proposed Wireless HDLC network is

$$E_C = N_T \left[P_{LO}\left(t_{ri} + t_{tx} + t_{st}\right) + P_{PA}\left(t_{tx}\right)\right] + N_R \left[P_R\left(t_{ro} + t_{rx} + t_{st}\right)\right], \quad (3)$$

where $t_{ri} = 10\,\text{ms} - (t_{rx} + t_{st})$ is the idle time between the start of the receive slot until the start of transmit slot and t_{ro} is the receive offset which depends on the synchronization accuracy. In our implementation the t_{ro} has a fixed value of 10 μs to compensate the PTP offset shown in Figure 9.

Table 4 shows the complete energy consumption of the Wireless HDLC communication protocol for a number of interesting points.

(1) Sending sink-to-SN (a) and receiving SN-to-sink (b) traffic have the same incremental cost, but sink-to-SN traffic has a higher fixed cost associated with the PTP Sync message. This is exactly as expected.

(2) Sending sink-to-multiple SNs (c) and receiving multiple SNs-to-sink (d) traffic differ in their fixed costs and in incremental cost. Sending sink-to-multiple SNs traffic has a high fixed cost, also due to the cost of sending a PTP Sync messages. Sending multiple SNs-to-sink traffic has a high incremental cost due to the cost of sending each data with an HDLC header and flag.

(3) Sending and receiving multicast traffic (c, d) and sending and receiving point-to-point traffic (a, b) were expected to show the different incremental cost associated with the number of HDLC frames in superframe.

(4) Sending and receiving WirelessHART point-to-point traffic (e) and sending and receiving point-to-point traffic (a, b) have the same incremental cost but WirelessHART traffic has a higher fixed cost associated with the Data-Link packet (DLPDU) size [37].

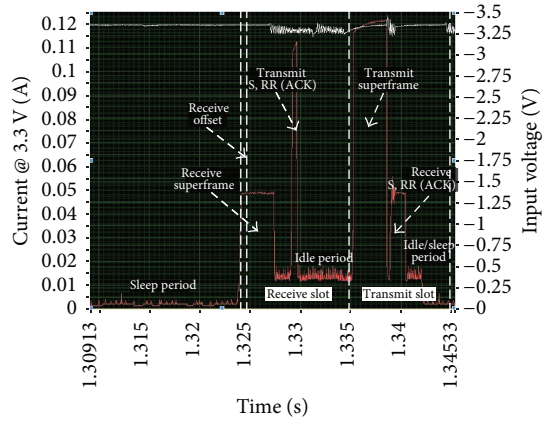


Figure 11: SN power consumption in hop-by-hop approach.

Each DLPDU consists of the following fields:

(i) 1-byte set to 0 × 41,

(ii) 1-byte address specifier,

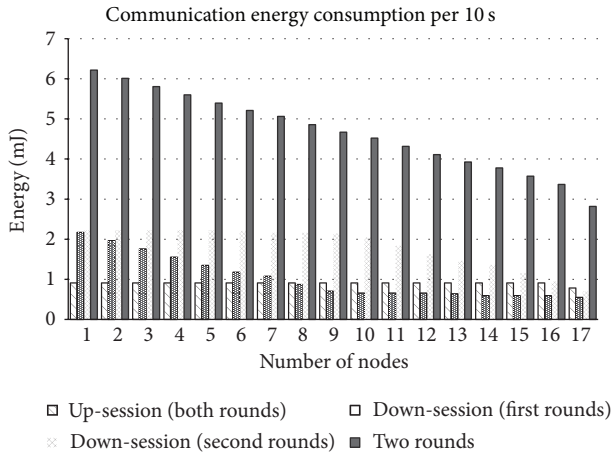


Figure 12: Communication energy consumption for the WUNS setup described in Section 4.4.

(iii) 1-byte sequence number,

(iv) 2-byte network ID,

(v) destination and source addresses either of which can be 2- or 8-bytes long,

(vi) 1-byte DLPDU specifier,

(vii) 4-byte keyed Message Integrity Code (MIC), and

(viii) 2-byte ITU-T CRC16.

Compared to WirelessHART, which has one of the lowest energy consumption between the standard WSNs, the Wireless HDLC has a lower cost, associated with the header size of the data link layer, which is crucial in large-scale WSNs. Generally, because ZigBee and 6LoWPAN networks are not using TDMA-based MAC, they have higher energy consumption and so they are not suitable for large-scale linear WSNs [18].

Figure 12 shows the communication energy consumption for the network setup described in Section 4. This figure illustrates the energy consumption for two communication rounds in order to obtain data from all the 17 SNs. Moreover, it provides energy consumption for each query from sink and for the SNs responses. Sending sink-to-SNs commands has a high fixed cost due to the use of multicast commands for the different types of SNs. Receiving data from the first group of SNs and the alert sensors has a lower cost for the last SNs in the linear network which maximize the idle period in the TDMA slots. Receiving data from the second group of SNs and the alert sensors has a fixed cost for the first SNs and an incremental cost for the last SNs based on the total size of the superframe. The higher energy consumption for the SNs situated close to the sink is exactly as expected, because of the routing functionality of these nodes and the size of the superframe when it gets close to the sink.

From simulations based on MATLAB, the number of nodes that can be connected to a Wireless HDLC network with data rates depending on the TDMA technique for constant data bits is depicted in Figure 13. The maximum data bit rate of approximately 100 kbps is obtained with one node

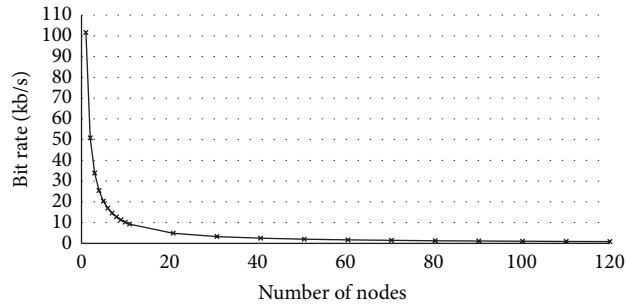


Figure 13: Maximum rate for aggregator per nodes with constant node bit rate.

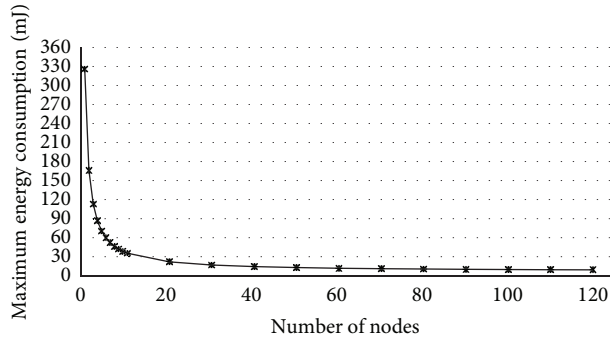


Figure 14: Maximum energy consumption per node for maximum sampling rate and maximum node bit rate.

connected to the sink and is decreasing with the number of nodes in the linear network to approximately 2 kbps for 50 nodes and 0,89 kbps for 120 nodes. Hence, for a linear network with devices generating data as in the evaluation network (80 bits/measurement), the data bit rate of 0,89 kbps for 120 nodes allows the system to sample all the nodes in approximately 10 seconds. The maximum energy consumption per node and the energy consumption per data bit in relation to the number of nodes in the linear network are depicted in Figures 14 and 15. As illustrated in Figure 14, for maximum sampling rate of 20 ms and maximum data bit rate of 100 kbps, the maximum energy consumption per node is approximately 330 mJ when only one node is connected to the sink. Therefore, in this configuration, the node will stay most of the time in active mode (receiving and transmitting data from/to sink node). With the increase of the number of nodes connected to the network, the total energy consumption per node decreases because the nodes will spend more time in sleep. However, the energy consumption per data bit increases with the number of nodes as shown in Figure 15 because each node spends more time in routing activities.

5. Conclusion

In this paper, the HDLC standard protocol implemented for low-power linear WSNs was presented and evaluated. This protocol is evaluated in order to meet the objectives of interoperability, efficiency, and reliability in WSNs. First

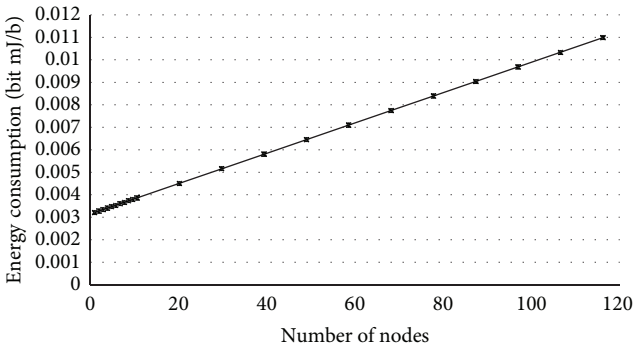


Figure 15: Energy consumption for bits per packet transmitted.

we introduced the large-scale infrastructure framework, focusing on the use of self-powered WSNs for protection and monitoring, the network architecture, and related works. Then we briefly presented the HDLC standard as a candidate protocol for linear wireless networks with bidirectional communication. We specified the advantages of this protocol in linear topology such as the possibility to transmit a variety of information, the use of poll/final procedure for control, or the use of unicast and multicast addresses with extending capability. We also proposed a novel way to create a half-duplex communication in a linear network based on low-power devices.

A hardware prototype for self-powered SNs, based on XBee PRO modules, has been developed for a large-scale infrastructure monitoring system. The preliminary results of a Wireless HDLC demonstration network with these sensors are very encouraging and also revealing. Using the HDLC protocol a user can interact at any time with different nodes in the network and can collect various types of data from many sensors at one time, which is an important capability for a linear network. Another advantage for low-power WSNs is that the HDLC does not increase the size of messages significantly and it does not introduce complex headers or fields to sensors data. Therefore, the Wireless HDLC is a low-power WSN which can be supplied entirely using harvesting approaches.

The investigation performed in this paper was an attempt to make open standard linear WSNs with an ultimate aim of achieving a standard low-power wide monitoring system, an essential requirement for the next generation of WSNs.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was carried out within the Sistemas Inalámbricos Para la Extensión de Observatorios Submarinos (SINEOS) Project under Grant CTM2010-15459 and the ENE2012-38970-C04-02 Analisis de Datos Basados en Aprendizaje automático y Sistemas Inteligentes de Adquisición de Datos Project under Grant CTM2009-08867 supported by the Spanish Science and Innovation Ministry (MINECO).

References

[1] S. Kim, S. Pakzad, D. Culler et al., "Health monitoring of civil infrastructures using wireless sensor networks," in *Proceedings of the 6th International Symposium on Information Processing in Sensor Networks (IPSN '07)*, pp. 254–263, Cambridge, UK, April 2007.

[2] I. Jawhar, N. Mohamed, and K. Shuaib, "A framework for pipeline infrastructure monitoring using wireless sensor networks," in *Proceedings of the 2007 Wireless Telecommunications Symposium (WTS '07)*, pp. 1–7, Pomona, Calif, USA, April 2007.

[3] B. Sheng and W. Zhou, "Ultra-low power wireless-online-monitoring platform for transmission line in smart grid," in *Proceedings of the 2010 International Conference on High Voltage Engineering and Application (ICHVE '10)*, pp. 244–247, New Orleans, La, USA, October 2010.

[4] D. De Caneva and P. L. Montessoro, "A synchronous and deterministic MAC protocol for wireless communications on linear topologies," *International Journal of Communications, Network and System Sciences*, vol. 3, no. 12, pp. 925–933, 2010.

[5] A. R. Devidas and M. V. Ramesh, "Wireless smart grid design for monitoring and optimizing electric transmission in India," in *Proceedings of the 4th International Conference on Sensor Technologies and Applications (SENSORCOMM '10)*, pp. 637–640, Venice, Italy, July 2010.

[6] Y. Yang, F. Lambert, and D. Divan, "A survey on technologies for implementing sensor networks for power delivery systems," in *Proceedings of the 2007 IEEE Power Engineering Society General Meeting (PES '07)*, pp. 1–8, Tampa, Fla, USA, June 2007.

[7] R. Liu, I. J. Wassell, and K. Soga, "Relay node placement for wireless sensor networks deployed in tunnels," in *Proceedings of the 6th Annual IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '2010)*, pp. 144–150, October 2010.

[8] "IEEE Standard for information technology—telecommunications and information exchange between systems—local and metropolitan area networks—specific requirements. Part 15.4: wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for Low Rate Wireless Personal Area Networks (LR-WPANs)," IEEE 802.15.4-2006, 2006.

[9] "IEEE standard for information technology—telecommunications and information exchange between systems—local and metropolitan area networks—specific requirement. Part 15.4: wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for Low-Rate Wireless Personal Area Networks (WPANS)," IEEE Std. 802.15.4a-2007 (Amendment to IEEE Std. 802.15.4-2006), 2007.

[10] I. F. Akyildiz and M. C. Vuran, *Wireless Sensor Networks*, Advanced Texts in Communications and Networking, John Wiley & Sons, 2010.

[11] ZigBee Alliance, http://www.zigbee.org/.

[12] J. Song, S. Han, A. K. Mok et al., "WirelessHART: applying wireless technology in real-time industrial process control," in *Proceedings of the 14th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS '08)*, pp. 377–386, St. Louis, Mont, USA, April 2008.

[13] IPv6 over low power WPAN working group, http://tools.ietf.org/wg/6lowpan/.

[14] "ISA100, Wireless Systems for Automation," http://www.isa.org/MSTemplate.cfm?MicrositeID=1134&CommitteeID=6891.

[15] P. Stig and C. Simon, "A survey of wireless sensor networks for industrial applications," in *The Industrial Electonics Handbook*, Industrial Communication Systems, chapter 12, pp. 1–10, CRC Press, 2nd edition, 2011.

[16] A. Calveras and A. Ludovici, "Implementation and evaluation of multi-hop routing in 6LoWPAN," in *IX Jornadas de Ingeniería Telemática (JITEL '10)*, pp. 1–6, Universidad de Valladolid, Valladolid, Spain, 2010.

[17] International Organization for Standardization, "Information technology—telecommunications and information exchange between systems—High-Level Data Link Control (HDLC) procedures," ISO/IEC 13239:2002, 2007, http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=37010&ICS1=35&ICS2=100&ICS3=20&showrevision=y&scopelist=CATALOGUE.

[18] S. Ganeriwal, R. Kumar, and M. B. Srivastava, "Timing-sync protocol for sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys '03)*, pp. 138–149, November 2003.

[19] "Standard for a precision clock synchronization protocol for networked measurement and control systems," IEEE STD 1588-2008, IEEE Instrumentation and Measurement Society, TC-9, The Institute of Electrical and Electronics Engineers, New York, NY, USA, 2008.

[20] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," *Ad Hoc Networks*, vol. 3, no. 3, pp. 325–349, 2005.

[21] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–105, 2002.

[22] J. Edgar, H. Callaway, and E. H. Callaway, *Wireless Sensor Networks: Architectures and Protocols*, CRC Press, 2003.

[23] R. B. J. A. Gutierrez and E. H. Callaway, *IEEE 802.15.4 Low-Rate Wireless Personal Area Networks: Enabling Wireless Sensor Networks*, IEEE, 2003.

[24] X. Bai, S. Kuma, D. Xua, Z. Yun, and T. H. La, "Deploying wireless sensors to achieve both coverage and connectivity," in *Proceedings of the 7th ACM International Sym-posium on Mobile Ad Hoc Networking and Computing*, pp. 131–142, Florence, Italy, May 2006.

[25] M. Zimmerling, W. Dargie, and J. M. Reason, "Energy-efficient routing in linear wireless sensor networks," in *Proceedings of the 2007 IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS '07)*, pp. 1–3, October 2007.

[26] T. Karveli, K. Voulgaris, M. Ghavami, and A. H. Aghvami, "A collision-free scheduling scheme for sensor networks arranged in linear topologies and using directional antennas," in *Proceedings of the 2nd International Conference on Sensor Technologies and Applications (SENSORCOMM '08)*, pp. 18–22, Cap Esterel, France, August 2008.

[27] U.S. Department of Commerce, "NIST framework and roadmap for smart grid interoperability standards," release 1.0, 2010.

[28] S. K. Singh, M. P. Singh, and D. K. Singh, "Routing protocols in wireless sensor networks: a survey," *International Journal of Computer Science & Engineering Survey*, vol. 1, no. 2, pp. 474–480, 2010.

[29] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information system," in *Proceedings of the IEEE Aerospace Conference*, vol. 3, pp. 1125–1130, Big Sky, Mont, USA, March 2002.

[30] *IEEE Standard for a Smart Transducer Interface for Sensors and Actuators Wireless Communication Protocols and Transducer Electronic Data Sheet (TEDS) Formats*, IEEE 1451.5, IEEE Instrumentation and Measurement Society, 2007.

[31] *IEEE Standard for a Smart Transducer Interface for Sensors and Actuators—Common Functions, Communication Protocols, and Transducer Electronic Data Sheet (TEDS) Formats*, IEEE 1451.0, IEEE Instrumentation and Measurement Society, 2007.

[32] "XBee and XBee PRO datasheet," http://www.digi.com/pdf/ds_xbeemultipointmodules.pdf.

[33] "MC13211 platform," http://www.freescale.com/webapp/sps/site/prod_summary.jsp?code=MC13211&nodeId=0106B9837F7103.

[34] D. M. Toma, J. del Rio, and A. Manuel-Lazaro, "Self-powered high-rate wireless sensor network for underground high voltage power lines," in *Proceedings of the 2012 IEEE International Instrumentation and Measurement Technology Conference (I2MTC '12)*, pp. 1881–1885, Graz, Austria, May 2012.

[35] J. Gao and L. Zhang, "Load-balanced short-path routing in wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 4, pp. 377–388, 2006.

[36] D. M. Toma, J. del Rio, A. Mànuel, and S. Gomariz, "Precision timing in TDMA-based Wireless Sensor Network through IEEE 1588 standard," in *Proceedings of the 19th IMEKO TC-4 Symposium and 17th IWADC Workshop Advances in Instrumentation and Sensors Interoperability*, pp. 683–686, Barcelona, Spain, July 2013.

[37] E. Shih, S. H. Cho, N. Ickes et al., "Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom '01)*, pp. 272–286, Rome, Italy, July 2001.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 158252, 18 pages
http://dx.doi.org/10.1155/2014/158252

Research Article

Protocol and Architecture to Bring Things into Internet of Things

Ángel Asensio, Álvaro Marco, Rubén Blasco, and Roberto Casas

Aragon Institute of Research, University Zaragoza, 50018 Zaragoza, Spain

Correspondence should be addressed to Álvaro Marco; amarco@unizar.es

Received 1 December 2013; Revised 16 February 2014; Accepted 16 February 2014; Published 13 April 2014

Academic Editor: Zuqing Zhu

The Internet of Things (IoT) concept proposes that everyday objects are globally accessible from the Internet and integrate into new services having a remarkable impact on our society. Opposite to Internet world, *things* usually belong to resource-challenged environments where energy, data throughput, and computing resources are scarce. Building upon existing standards in the field such as IEEE1451 and ZigBee and rooted in context semantics, this paper proposes CTP (Communication Things Protocol) as a protocol specification to allow interoperability among *things* with different communication standards as well as simplicity and functionality to build IoT systems. Also, this paper proposes the use of the IoT gateway as a fundamental component in IoT architectures to provide seamless connectivity and interoperability among *things* and connect two different worlds to build the IoT: the *Things world* and the *Internet world*. Both CTP and IoT gateway constitute a middleware content-centric architecture presented as the mechanism to achieve a balance between the intrinsic limitations of *things* in the physical world and what is required from them in the virtual world. Said middleware content-centric architecture is implemented within the frame of two European projects targeting smart environments and proving said CTP's objectives in real scenarios.

1. Introduction

Since the last decade, we are assisting in a progressive jump from a nonubiquitous Internet, where humans access Internet using a computer at their work or at home, to the current ubiquitous Internet where we access the Internet using smartphones, tabs, or TVs, anytime, anywhere. In the same way, now comes the time of the Internet of Things (IoT) when not only humans but also *things,* any object surrounding us, are present in Internet [1].

Things must be considered in the broadest sense of the word as real or virtual entities that exist and evolve in a context and time and have univocal identifiers. On the other hand, the term Internet applied to them conveys the idea that all these *things* are heavily communicated and interrelated among them. Commonly IoT can be approached from different perspectives: *Internet* for communications, cloud, and services; *things* for physical elements, sensor networks, and user interfaces; and *semantic* that considers ontology of *things* in Internet [2].

Ideally, *things* on the IoT will have full interconnectivity and computation resources, being natural to consider connecting these *things* to the web using current paradigms. Different works investigate the types of *things* (sometimes called smart objects), their nature, and relationship with the IoT; according to the different perspectives, a *thing* has awareness, representation, interaction, and so forth [3]. The classification of *things* into standard groups helps to show that they are the physical part of IoT with constraints and needs that must be taken into account. So, coming down to implementation, while IoT concept talks about ubiquitous, invisible, and context aware *things*, technology poses hurdles such as energy supply, price and size of devices, seamless connectivity, or interoperability [4]. Additionally, as, currently, Internet has a wired backbone, "being there" forces all the *things* to be IP compatible and use access points or gateways to bridge global fiber optic or cabled infrastructure.

Wireless communication protocols are mandatory if *things* need to be mobile and ubiquitous. Depending on the specific application, and leaving aside proprietary protocols,

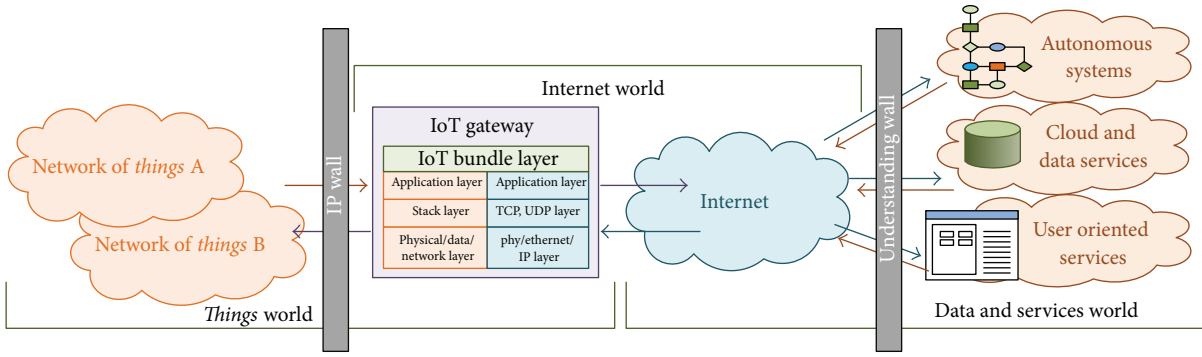


Figure 1: IoT architecture.

different standards are commonly used to provide connectivity: ZigBee, RFID, Bluetooth, 6lowPAN, WIFI, 3G, and so forth. Many works evidence the importance of selecting the most appropriated technology in each situation comparing them in terms of network topology, coverage, data throughput, or energy consumption. In any case, the more data you need to exchange, the further you need to communicate, the more time you need to be online, then more energy you need, and consequently the quicker you will run out of batteries.

The arrival of IoT will lead to appearing multitude of new services, improving the quality of life of people, and offering new business opportunities. Ultimately, the IoT is expected to bring a revolution in the concept of society, similar to how Internet changed the concept of communications and information [5]. In the last few years, initiatives and project related with IoT have been growing in number and scope throughout the entire world. The IoT concept is becoming tremendously popular and already appearing systems that claim to offer IoT solutions to the final user without high technical requirements [6]. Current developments are in early stages being most of the services based on monitoring or sensing variables to extract information and then analyze and represent them [7].

It is considered that the development of IoT will play an important role in the near future; thus, public and private investments have been made in R&D, demonstration, and deployment activities [8, 9]. To date, most of IoT solutions are small subnets of interconnected objects. It is not possible to talk about IoT until all objects are interconnected and to improve this interconnectivity, an architecture that ensures interoperability between systems is mandatory. Thus, both public and private initiatives are currently focusing on standardization of the IoT [10–14].

In summary, the evolution of IoT implies overcoming real *things*' limitations enabling them to communicate with the Internet using a common language. In this paper, we propose CTP (Communication *Things* Protocol) as an ontology-based solution to enable understanding among *things* and the use of an IoT gateway to take *things* to the Internet world.

2. Internet of Things

2.1. Architecture. Most IoT implementations follow an architecture that contains different worlds each of them with their own characteristics; Figure 1 shows an example [15].

The *things* world relates to michroelectromechanical systems, smart sensors, simple human-machine interfaces (HMIs), and so forth, which associate in networks (Networks of *Things*, NoT) to ubiquitously interact with other *things*, the environment and/or people. NoT are usually resource-challenged ecosystems, typically with medium-high time access delays, high error rates, low data throughput, and limited online time where energy consumption must be optimized to the maximum. In a simplified model of a *thing*, basic blocks of communication, computation, and interaction (sensors, actuators, and HMI) are distinguished.

The Internet world usually is constructed around computers, centralized software infrastructures, or in the cloud [16]. Applications and services use *things* to provide context awareness, artificial intelligence, affective computing, and so forth [16]. Depending on their consideration, tablets and smartphones can be included in both worlds. Nevertheless, due to their power (computing, communications, battery lifetime, user interfaces, etc.) we find it more appropriate to consider them closer to the world of computers and Internet than to the world of *things*. Currently, Internet acts as the base infrastructure for the exchange of information. However, access to it has several constraints such as the need for unique identifier, the change of communication technology, and the adoption of the Internet Protocols. Nowadays, this "IP wall" is usually jumped through an IoT gateway.

Internet is the global interconnection method where multitude of services and applications use extracted information of IoT to provide services to end users. On each of them, needs are a virtual representation of the *things* of the IoT in order to enable interaction. Once "IP wall" is saved and thanks to the connectivity provided by the Internet, services can access the information, but, for the information to be useful, it must be understood and this poses what we call the "understanding wall."

2.2. Interoperability. IoT interoperability implies capacity to both exchange data (crossing the IP wall) and understands the information embedded in data (crossing the understanding wall). Communication standards ensure stack layer's communication between *things* in the same network sharing the same protocol, that is, network management and maintenance, security, data exchange, and so forth. Nevertheless, if application layer is not defined, *things* will not understand among them unless previously agreed between developers, that is, information understanding. This is the case of 6lowPAN, RFID, WiFi, or cellular; for example, if two manufacturers want to develop 6lowPAN temperature sensors and thermostats are able to interoperate among them, there is no common definition to adopt; they will need to agree on the protocol exchange application-related information, temperature data format, procedure to virtually bind devices, and so forth.

Automation and control networks such as Lonworks, BACnet, Konnex, or CANOpen define how devices are represented in their networks; objects and variables are the most common approaches. Bluetooth and ZigBee go one step further in interoperability defining profiles and device objects within the application layer. Bluetooth defines profiles (hands-free, health device, human-interface device, etc.) corresponding to vertical applications. For example, we could build a monitoring infrastructure with Bluetooth microphones streaming audio according to hands-free profile; no matter their manufacturer, any certified host compliant with the profile will play the audio gateway role without any additional programming [17].

ZigBee not only defines vertical profiles (home automation, energy metering, healthcare, etc.), but also defines horizontal functional domains to specify how devices must exchange application data attending their functionality [18]. For example, every ZigBee compliant temperature sensor must implement "measurement & sensing functional domain" and any other device in the network (e.g., a thermostat) would be able to get temperature information as defined in the specification. Additionally, ZigBee allows instantiation of intelligence in the network by defining how to coordinate devices to produce scenes and create virtual bindings among devices [19], for example, to program a switch to turn on two lights and open a motorized door.

Sometimes two specifications coordinate to solve interoperability, such as ZigBee that specifies how to interoperate with BACnet. Devices using any wireless communication standard (e.g., ZigBee) cannot directly interoperate with devices over other standards (e.g., WiFi) unless there is a protocol aggregator and translator connecting both worlds, the IoT gateway.

Although IP connectivity is not necessarily required to ensure inter-*thing* communication, many standards include IP as part of their specification to ease the process to connect *things* to the Internet. Nevertheless, this is not enough to ensure interoperability at required IoT application level. As *things* are resource-challenged communication nodes, efficient data transmission mechanisms are needed at IP level; CoAP [20], XMPP [21], RESTful HTTP [22], and MQTT are relevant specifications at this level.

Once efficient connectivity between devices is granted, standards such as EEML (Extended Environments Markup Language), SensorWeb (including SensorML and TransducerML), or SenseWeb provide interpretation to bytes exchanged integrating sensors and actuators with the virtual world. These schemes just express queries and data modeling, lacking of semantics, and ontologies that are necessary for complex information processing and support to service composition and adaptation at higher levels of abstraction.

IEEE1451 standard is a network-independent specification for smart transducers (sensors or actuators) that provides a common language regardless of the protocol used. The standard defines different application profiles: environmental (climate monitoring, greenhouse gases, and other chemical sensors), smart meter (monitor water, gas, or electricity consumption), health care (monitor the body using external or implantable sensors), and smart home automation and industrial (pipe's monitoring, comfort, and surveillance sensors) [23]. The standard is based on the Transducer Electronic Data Sheet (TEDS) to describe a set of communication interfaces for connecting transducers to microprocessors, instrumentation systems, and control/field networks. According to IEEE1451.0 specification, TEDS provides information about transducer's identification, operation, calibration, manufacturer, and so forth.

Lying on IEEE1451.5 specification for radio-specific protocols [24], it is possible to implement wireless sensor networks using IEEE1451 over communication standard protocols such as Wifi [25], Bluetooth, ZigBee [26], or 6LowPan [27]. Similarly, IEEE1451.2 defines cabled SPI and UART, IEEE1451.6 over CANOpen, and IEEE1451.7 over RFID. This interoperability is enabled by the Network Capable Application Processor (NCAP) that aggregates the different Transducer Interface Module's communication standards (TIM) over the same common language. Through NCAP and following IEEE1451 it is possible to implement web services that make *things* discoverable, accessible, and controllable via the Internet 22. In the same line, ZigBee defines the ZigBeeGateway Public Application Profile that specifies how devices must be connected to the Internet and to service providers.

Initiatives, such as Sensei project [7], COSE [28], DogOnt [29], and other [30] different ontological approaches to model *things* (sensors, actuators, simple human-machine interfaces, appliances, etc.) in smart environments, associate contained devices through semantic relationships [31] and seamlessly integrate *things* with web services.

3. Common Things Protocol (CTP)

3.1. Rationale. Networks of *things* (NoT), defined as smart infrastructures of embedded devices with local intelligence and access to the "information ether," are the base of the IoT. As a part of the Internet, one of the basic needs is the interaction mechanism between agents which implies that, between them, there must be connectivity and understanding in order to provide interoperability. To achieve this goal, different IoT protocols are being proposed to provide efficient,

seamless, and robust connectivity (CoAP, XMPP, RESTful HTTP, MQTT, etc.) but not providing semantics. From an integral perspective, IEEE1451 appears to be a good suited option for the IoT as it tackles specification from sensor to network interface providing independence from the communication protocol, enabling self-identification of devices, long-term self-documentation, plug and play capacity to ease field installation, upgrade, and maintenance [32]. On the other hand, the standard has a limited penetration in IoT applications out of the electronic instrumentation field. Reasons for that can be the little compatible hardware available (no commercial wireless sensor networks consider it) because IoT developers mainly come from the computer science field (usually considering semantic approaches), due to complexity of the standard [33], or the excessive detail in electronic aspects that are not commonly used by the application (e.g., transducer channel reads delay time or incoming propagation delay through the data transport logic). Communication standards such as ZigBee or 6LoWPAN are much more used in IoT applications but involve hardware restrictions and lack of interoperability among them.

The Common Things Protocol (CTP) aims to provide a specification that allows interoperability among communication standards and coexistence with IoT protocols, but prioritizing simplicity, efficiency, and functionality to build final IoT systems.

CTP takes into consideration existing specifications in the standards and the needs from the final applications that are to be supported by the *things* in the field. It integrates the strategies, concepts and terms of some of the alternatives, for example, IEEE1451 TED's concept to provide device's information, sensor and actuator working modes, and so forth, or ZigBee concepts of clusters and endpoints. CTP definition is approached from an ontological perspective considering *things*, not just as sensors [34] but as electronic devices that perform some sort of function in the IoT application being in contact with user and context and usually fulfilling the following paradigms:

(i) *context interaction:* embedded sensors (to sense the user/context), actuators (to modify the environment), and/or simple human interfaces (to interact with people);

(ii) *computin:* to have computing capabilities and memory that allow them to implement from the simplest logic to complicated services or data processing algorithms;

(iii) *communication:* to have at least one, usually wireless, communication media commonly following a standard and adapted to communication requirements (range, power consumption, and data throughput). It is required to interoperate among them and integrate in the Internet. Thus, unless they are able to directly connect to Internet, a network gateway is required to allow interoperability among different networks of *things* and to provide Internet access;

(iv) being an *electronic thing*: as anything in this world, regardless it is electronic or not, each *thing* is unique and "lives" in a specific space and time. Additionally as any electronics, it needs energy to operate.

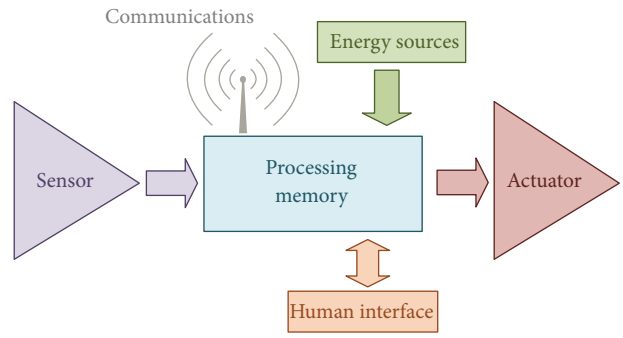


Figure 2: *Thing* block diagram.

Figure 2 represents said paradigms in the block diagram of a *thing* able to integrate in the IoT.

Thus, derived from their capacity to interact with the context, CTP considers that *things* can have three different functionalities:

(i) sensors that gather and process information from the real world in environmental and person-centric contexts [35]. Information from the environmental context is useful to create an objective snapshot of what is happening in every moment, while person-centric helps to understand and evaluate the objective data in order to extract conclusions, to define guidelines, or to identify patterns adapted to each user;

(ii) actuators that provide the ability to act over the environment. IoT applications may need to act over the environment when the user is not able to control (e.g., because he/she has a disability) or he/she is not aware of specific situations that require actuation (e.g., forget about the heating when going to sleep) or simply he/she is not present in the environment;

(iii) pervasive human-machine interfaces including traditional (switches and dimmers) and new interaction paradigms, providing the user with relevant information or notifying him/her about events in new and natural ways (e.g., a color device turns red when house's energy consumption is high).

Regardless its functionality, each *thing* will always "live" in a certain location and time and will have a processor, communication transceiver, and power source. Similar to MetaTEDS in IEEE1451 and ZigBeeDevice Object (ZDO) in ZigBee, this constitutes the basic set of attributes and functionalities of any device, which is modeled by the endpoint BASE in CTP. Depending on the specific functionalities it integrates, it will also implement multiple application endpoints that should be of any of the said categories: SENSOR, ACTUATOR, or HMI.

Thus, an endpoint can be defined as each of the subdevices that have a complete functionality and together with others build the *thing*; in total we just define the said 4 endpoints to model any*thing*. Additionally, a cluster is defined

as a set of commands, events, and responses (some mandatory to implement in order to guarantee interoperability) which together define a communication interface between two endpoints. Clusters constitute the implementation of the ontological representation of *thing*'s nature; for example, endpoint BASE includes location, time, and power clusters. Endpoint and cluster concepts are shared with ZigBee and are similar to transducer channels in IEEE1451.

Commands are usually action requests to an endpoint (of the same or different *thing*) that should send back a response informing about the action result, for example, ask for a sensor value and get it back. The ontology defines attributes which are implemented in the protocol as GET/SET requests and responses. Events are asynchronously generated messages sent to previously subscribed endpoints; for example, presence detected by a sensor is sent to light actuator. When defining a communication interface, two strategies can be adopted: (i) using a small, but well defined, number of messages, where the versatility is achieved by parameters (ii) or defining a greater number of messages allowing more specific control with lower parametric content [36]. An example of this duality is to define a single configuration command with many parameters, or several commands to adjust each parameter independently. The selection of one or another philosophy conditions ease of use, generalization capacity, adaptability to different scenarios and future maintenance, and backwards compatibility. CTP chooses to define a reduced and simple communication interfaces defining the meaning type and range of the parameters when needed. For example, many clusters have in common a read-only information attribute; in the case of sensors it provides information of ranges, accuracy, format, units, and so forth of the measure it provides, while for actuators the same attribute indicates how the actuator should be controlled. Similarly, some clusters include a read-and-write configuration attribute, whose specific parametrization depends on the device.

3.2. Endpoints and Clusters. Figure 3 shows the four endpoint types (BASE, SENSOR, ACTUATOR, and HMI) with their associated clusters and the main attributes, commands, responses, and events.

3.2.1. Endpoint BASE. All devices, regardless their nature, must have endpoint BASE containing the generic and common characteristics whatever their functions are. Related with this endpoint we define the following clusters.

Cluster DEVICE. It manages *device identification*, *description* of their functionalities (as in IEEE1451 TEDS's globally unique identifier and transducer channels), and minimum *self-operation* functionalities (as BIOS). Its implementation is mandatory in all CTP devices, as it is the basis for ensuring interaction with other system elements. Devices may have the ability to operate in different ways depending on the context; the *mode* parameter is used to switch between them and to facilitate the use of the *thing*, by providing to a nonadvanced user a set of predefined modes of operation that do not need any previous settings to set mode and operate. Also, this cluster has several capabilities oriented to work with low power devices, for example, and similar with TEDS, a *timeout* parameter that indicates the amount of time after an action for which the lack of reply following the receipt of a command may be interpreted as a failed operation.

Cluster LOCATION. Each device will always have a location that can be essential information for many services. It can be known or not, it can change or remain, and device can self-calculate it or be written externally, whichever the case, this cluster allows getting and setting *device's position*, programming its timed update, or reporting it in response to specific events.

Cluster POWER. As location, every device will have a *power source* (battery, mains, energy harvesting, etc.); its type, consumption profile, and energy remaining are the main information deal within this cluster.

Cluster TIME. Again as location and power, every device "live" in a specific instant of time, and whether relative or absolute this cluster allows the management of temporal information such as time synchronization and scheduling of timed actions.

Cluster PROXY. Similar to TEDS, this cluster acts as an aggregator of endpoints (or transducer channels) and datalogger manager (if memory is available) simplifying access to data. It reduces communication burden and thus increases battery efficiency.

Often, implementation of networks of smart sensors and actuators goes through a central device that receives data from sensors, processes the information, and then orders actions to the actuators. This has a number of problems: (i) density of data traffic, (ii) increased energy consumption, (iii) masking the mesh concept, since there is a central node acting as sink and source of information, and (iv) reducing robustness to failure. To face that, it is possible to define a basic distributed intelligence by subordinating the behavior of some devices to the events generated by other devices.

Cluster BEHAVIOR. Similar to ZigBee this allows the creation of groups of devices or endpoints and the establishment of logical relationships (bindings) among them. We expand its native definition defining the concept of a trigger event for a *binding* which allows that any endpoint generating events (e.g., timing, threshold event, etc.) could trigger any endpoint(s) in one or more devices with their corresponding parameters (e.g., changing the mode to low power of a device). Also, as in TEDS when defining embedded actuators, these bindings can be internal to a device and serve, for example, to trigger actions; button event triggers a light actuator.

This cluster also allows delegation of system's intelligence in the devices as it provides methods to program autonomous operation based on the specification of operational rules. It is based on logical rules that verify compliance with certain conditions of different endpoints of the device. Each situation

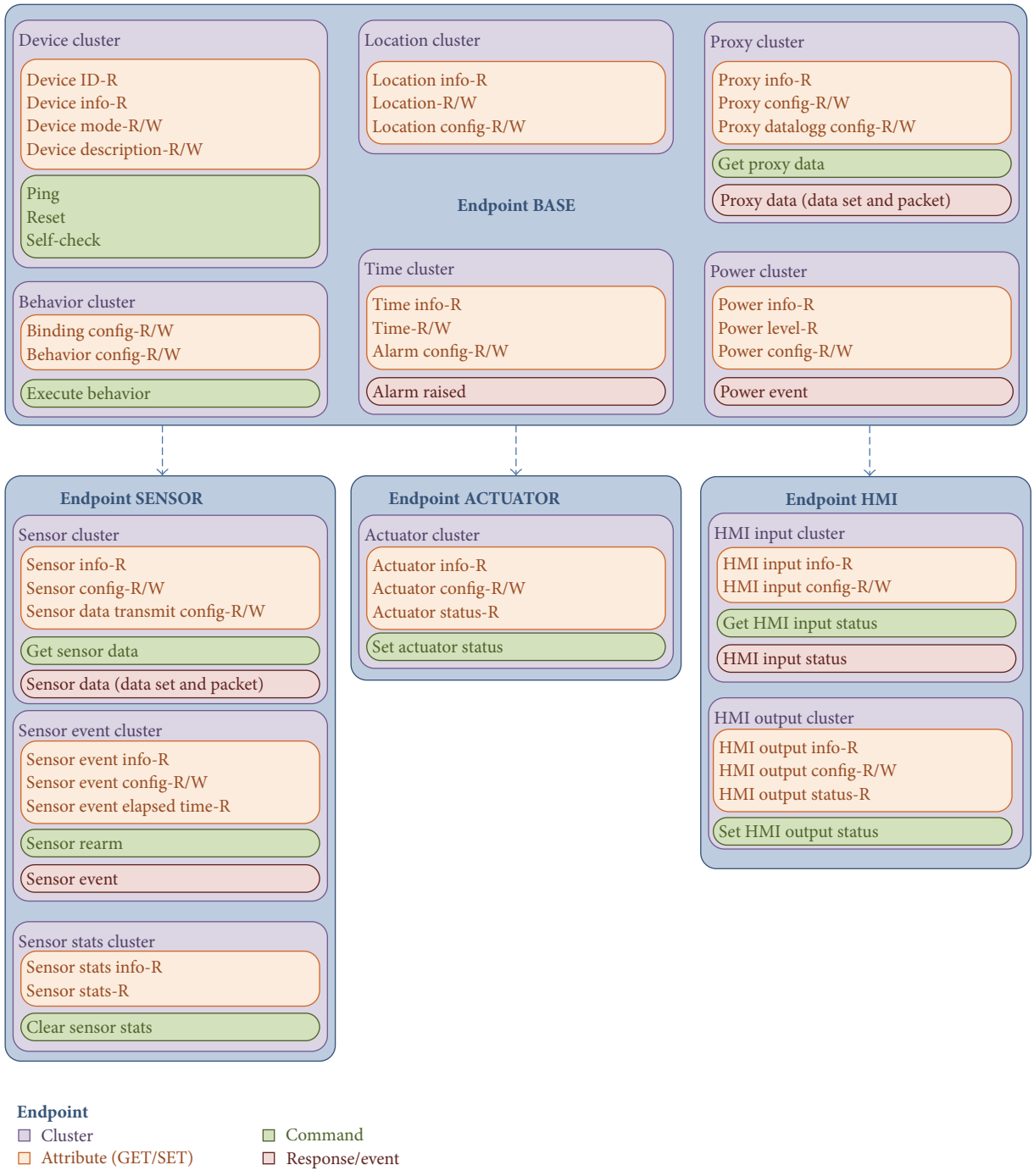


Figure 3: Summary of endpoints, clusters, and main interfaces in CTP.

that activates a behavior is called activation scenario. Notice that as a result of a rule of behavior a binding could be triggered.

3.2.2. Endpoint SENSOR. Sensors are elements that can extract information from the context. CTP allows management of basic features and supports more abstract and complex capabilities.

Cluster SENSOR. As transducer channels TEDS, it manages the basic actions to request measure, defines sensor settings (e.g., sampling period), and defines the mode for transmitting the data set or data packets aggregating several measurements (e.g., on command, timed, or buffer full). It also provides essential attributes of the measurement such as units, accuracy, data ranges, warm-up time, vectors, and data packets. Every sensor type must implement this cluster.

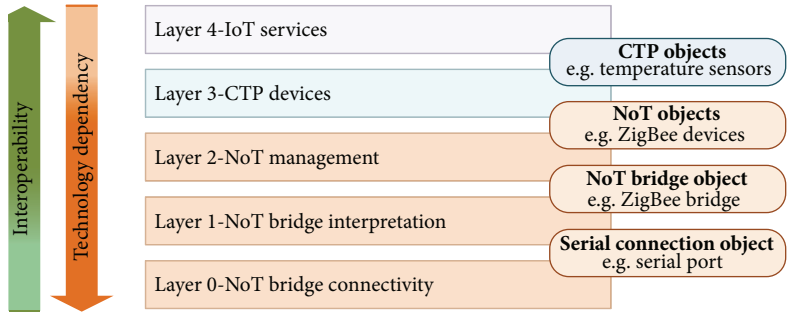


Figure 4: IoT gateway middleware architecture.

Cluster SENSOR_EVENT. Again as transducer channels TEDS it configures event triggers associated with a sensor (e.g., upper and lower thresholds, bit patterns, etc.) and provides the events.

Cluster SENSOR_STATS. This cluster performs local preprocessing and analysis of the captured data. This is of special interest when we are interested in obtaining a summary of the observation (e.g., maximum values, average values, etc.). As energy consumption between communication and computation in a wireless sensor node has a high ratio ("sending one bit" versus "computing one instruction") [37], its implementation is especially interesting in low power applications.

3.2.3. Endpoint ACTUATOR. While sensors allow obtaining contextual data, actuators are the elements that provide means to act over the environment. In relation with the IoT, an actuator can be considered as a device that transmits information or energy to another power mechanism or system (motors, electromagnets, thermocouples, heaters, coolers, etc.) that finally alters the environment. Thus, a *thing* with actuation capabilities is usually a device with control outputs, which is connected to an electromechanical system that opens doors, windows, shutters, control lighting, heating, and so forth.

Cluster ACTUATOR. It manages basic actions, controls the states of the actuator, and defines its configuration parameters. These states can be as simple as on/off or define complex operation patterns such as open the door for two minutes and then close and lock it.

3.2.4. Endpoint HMI. HMI (Human Machine Interface) devices are aimed to interact with a human user. In fact, they could be considered as sensors and actuators to interface people (in the same way that sensor and actuator connect with the environment). But given its importance and different use from the application perspective, it is convenient to assign its own type of endpoint. CTP considers the following clusters.

Cluster HMI_INPUT. It manages configuration and operation of simple input devices interface, such as buttons, switches, or dimmers. It also considers groups of elements such as arrays of keys to model a keypad.

Cluster HMI_OUTPUT. It manages basic operation of simple output devices such as LEDs and buzzers. It also considers groups of elements such as arrays of LEDS to model a LED strip.

4. IoT-Gateway Architecture

According to the proposed architecture, the different *things*, maybe on several NoTs, must interconnect to the Internet, that is, jumping to the IP world. In most cases, this involves changing not only the communications protocol but also communication technology, which imposes the need of a gateway interfacing *things* and Internet worlds.

In the proposed architecture, the gateway is not a single bridge between protocols; it is also an intelligence aggregator and distributor of services. The IoT gateway must support management (discovering, recruiting, and connecting devices) in the NoTs and provide interoperability between them. Such duty requires that the IoT gateway have enough power computation to handle requests and responses from both domains and a flexible architecture to ensure interoperability. To accomplish such task, the layered middleware architecture shown in Figure 4 is proposed.

Lower layer of the middleware is in charge of providing base connectivity with the NoT through the device acting as the network bridge, typically a USB dongle, a Bluetooth device, or a TCP socket which results in a sort of serial connection object allowing sending and receiving bytes as data streams.

Second layer adds meaning to those bytes, implementing the specific communication protocol of the bridge and allowing effective communication with the NoT. This results in an object representing the bridge, which encapsulates the communication protocol with appropriate methods and fields.

Middle layer is in charge of managing the NoT through the usage of the bridge representation service, being capable of discovering network devices (*things*), recruiting them, and handling network unavailability. At this layer, objects representing network devices are available, although they are already described in terms of the underlying technology.

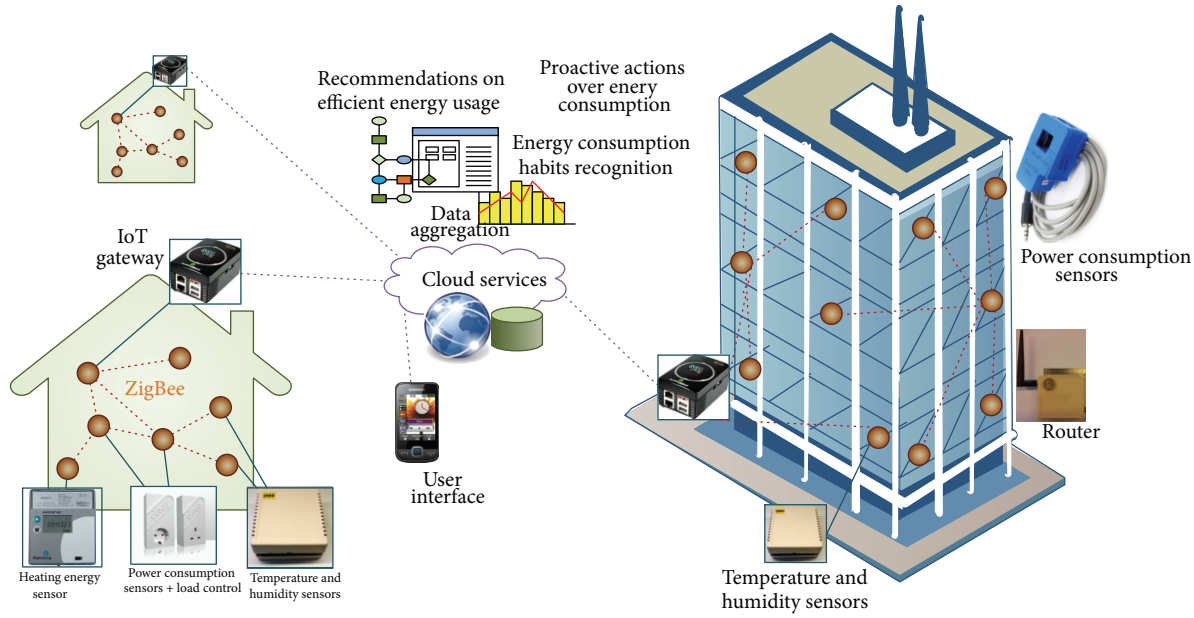


Figure 5: Test scenarios.

In the next layer, objects related to NoT devices are described as CTP objects, regardless their network technology, and full interoperability among *things* is achieved. CTP objects are described as a main device representing the endpoint BASE plus a collection of *channels* related to remainder endpoints.

Upper layer is devoted to gateway IoT services, which use CTP objects to link them to remote services (such as sending data to a remote database or allowing accessing a *thing* from a remote client) or build local services to perform offline operations.

5. Experimentation and Results

CTP has been successfully applied in several projects, but the largest implementations have been done in two European projects: "Easy Line Plus" [38] in the domain of Ambient Assisted Living and "Renaissance" [39] in the domain of Energy Efficient Buildings. The Easy Line Plus project sets the frame to define the CTP protocol, but it was within the Renaissance project where all the aspects described above have been formally deployed and tested. Therefore, we will describe below the Renasissance project setup.

5.1. Test Scenario. "Renaissance" main objective is energy saving, through the implementation of bioclimatic buildings, urban planning and rehabilitation, incorporating renewable energy, and reduction of energy consumption in households by improving habits of energy usage by users. Through remote data analysis, the system derives user patterns relating their energy consumption and comfort variables and then issues recommendations in order to increase their awareness and enhances energy savings. This requires an accurate and long-term monitoring of energy consumption and environmental conditions in order to generate enough data to extract relevant information.

To meet the needs of this project, the IoT paradigm is the ideal solution. The simplified structure of the system is a number of *things* that ubiquitously extract context information (temperature, humidity, heating energy, and power consumption). Through the proposed architecture, this information is handled (data aggregation, data base management, and cloud computing) and analyzed (habits recognition) to provide a range of services (data visualization, user profiling, assessment of the best practices, and user information through multimodal user interfaces).

Technically the system developed followed the architecture in Figure 1; *things* use ZigBee standard plus CTP; IoT gateway is a Linux embedded PC running the already presented middleware, and services run in different hosts (PC, mobile phones) using data stored in the cloud. As evaluation has been done in a real and operative deployment, the system previously ensured a number of quality requirements such as stability, ease of deployment and maintenance, and low intrusiveness in the context. Similarly, a number of limitations related to the *things* have been solved, namely, reducing power consumption to ensure months of operation with the same batteries, cost-effectiveness, and reduced size.

Two different types of installations have been done as follows (Figure 5):

(i) 80 m^2 dwellings with one ZigBee network formed by 3 ambient (temperature + humidity) sensors, 1 heating energy sensor, 1 monophasic power consumption sensor, and an IoT gateway;

(ii) 20.000 m^2 building with one ZigBee network formed by 70 ambient sensors, 20 triphasic power consumption sensors, 17 routers, and an IoT gateway.

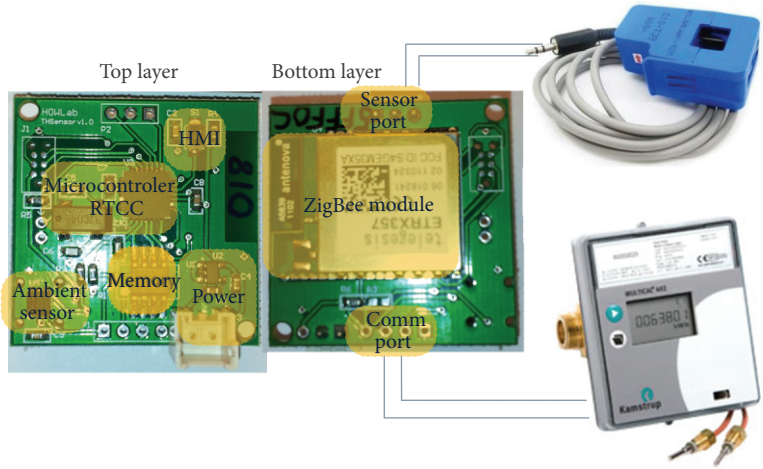


Figure 6: Zigbee *thing* hardware implementation.

In both cases, although the environment is quite different, the technological development is similar, allowing assessing and evaluating CTP in different scenarios.

The system has been installed and was running for 9 months in 43 dwellings simultaneously (i.e., 43 systems) in Zaragoza, proven to be stable, no maintenance was needed, and the expected functionality was provided. The building installation has been running for 6 months (already running) at the University of Zaragoza [40].

5.2. ZigBee Network of Things. Besides forming a ZigBee network, the sensors developed allow sensing the physical environment, connecting to analog or digital simple external sensors (e.g., presence detector, magnetic contact, etc.), and connecting to external commercial smart meters (e.g., commercial electricity or heating water consumption).

The aforementioned 20.000 m^2 building with long corridors, anti-fire doors, and so forth is a challenging scenario for wireless communication network with a hundred of nodes. A multihop mesh topology with an overlap of coverage areas and redundant communication paths was deployed. The ZigBee network backbone is formed by 17 routers, 1 network coordinator, and a data sink connected to the IoT gateway. The infrastructure is devoted to keep routing tables to date and interconnect 90 CTP *things* (70 ambient sensors and 20 power consumption sensors) that are ZigBee Sleepy End Devices. Any sensor enters low power mode between measurements, being disconnected from the network along this time, and its *father* (a router) holds its messages. Periodically, sensors poll their parents to check for incoming messages. This time between polls introduces latency in the communication but reduces energy consumption of the *thing*; to avoid possible loss of messages it is set to 4 seconds. Also, time between measurements, time between sending data, timestamping, and so forth can be configured in order to balance power consumption and application requirements. In this application, environmental nodes measure and immediately send data every 10 minutes, while power consumption sensors measure and store data every minute and then send it every 10 minutes. As it is not necessary that measures are simultaneous, the update process of the system is randomized to reduce the probability of several *things* trying to send messages at the same instant, which would increase medium access time and, therefore, energy consumption.

5.2.1. Hardware. Hardware implementation (Figure 6) is designed to be versatile. Device implementation is based in a dual hardware architecture where a low power microcontroller (PIC18F26J11) runs the main application and controls the network coprocessor that implements ZigBee Pro stack (Ember 357). After deep analysis and experimentation, we found architecture more efficient in terms of power consumption than using a system on chip solution (embedding a radio module plus a programmable microcontroller) as it allows splitting tasks between two specialized microcontrollers, one for sensing and processing tasks and the second for communication [41]. The device additionally has an EEPROM memory, a real time clock, expansion ports, a button, a LED, and a temperature and humidity digital I2C sensor (Sensirion SHT21). Along with these onboard capacities, the hardware features two expansion ports: first for connection of external analog/digital sensors and second for communications to external systems. Thanks to them, different *things* are built: a noninvasive AC current sensor (a SCT-013-000 0-100A split core current transformer) enables power consumption sensor and a communication port connected to a commercial device (Kamstrup) for measuring heating water consumption.

5.2.2. Firmware. Accordingly, the basic configuration of the device (only onboard capabilities) presents 5 endpoints: base (base type), temperature (sensor type), humidity (sensor type), button (HMI type), and LED (HMI type). Note that although SHT21 sensor is a single electronic component that measures temperature and humidity, there are two different endpoints one for each magnitude.

Nevertheless, similar to IEEE1451, access to both endpoints is possible using the proxy cluster within the base endpoint. According to the devices connected to the ports described above, news endpoints (power and heating water consumption, both sensor types) are added when necessary. The CTP protocol is therefore suited to the different characteristics of the hardware, while performing an abstraction of it.

Clusters implemented per endpoint are as follows:

(i) BASE: device, power (for battery level monitoring), time (to provide timestamps), proxy, and behavior (to program clima control when event happens),

(ii) SENSOR: sensor (to provide single measurements), sensor event (to get events when upper and lower thresholds are surpassed), and sensor stats (to provide maximum and minimum levels),

(iii) HMI: HMI input (to handle button events) and HMI output (to handle LED operation).

The use of CTP provides a structured and simple programming strategy, which results in modular code with high reusability. Figure 7 shows a simplified view of the proposed structure for a common firmware of a *thing*: configuration of the microcontroller, peripherals, ZigBee communications, and CTP and system behavior.

5.3. IoT Gateway. An embedded computer running Linux, where the middleware described before was implemented, acts as the IoT gateway. The middleware architecture falls within the SOA (Service Oriented Architecture) paradigm, and we use OSGi [42] (Open Services Gateway initiative) as development framework. OSGi defines a framework where pieces of code are organized into bundles that can be managed separately. OSGi bundles are agents which might be dedicated to specialized tasks, such as handling a serial port, providing a command line interface, collecting, aggregating and analyzing data, and so forth. These bundles communicate and interact with each other by means of services which are published within the framework, and each bundle can acquire and utilize them. The main strength of OSGi is that the framework manages these bundles dynamically, allowing them to be upgraded without terminating the full application, as well as enabling the availability of the services to other bundles depending on the situation. That allows us provinding new features and capabilities to the IoT Gateway by adding new services, which may use the services already existing in the framework, but keeping current features unaltered.

Using OSGi as development framework also eases development of the middleware layers, as we may focus on one single layer when developing, comprising one of more bundles, while the whole application layers will be arranged on execution time. Middle-layer interfaces are also defined to specify interoperability among adjacent layers and eliminate direct dependencies among bundles implementing each layer. According to Figure 8, the different layers have the following objectives.

(i) The network bridge is a USB dongle with a ZigBee module that can be controlled through an AT command interface on a serial port. On the lower layer, the Serial Communication middle interface defines a SerialConnection service, which basically allows writing bytes and notifies about incoming bytes, and a SerialDriver service, which is in charge of creating and discovering available SerialConnection services. Main bundle implementing this layer operates the UART interface, but other stream-based interfaces, such as a Bluetooth connection, may adopt the same middle-layer interface, allowing communication with Bluetooth devices transparently to the upper layers. In our particular case, we have implemented an application service that wraps a SerialConnection into a TCP socket, and at the remote client, another bundle creates a SerialConnection services that allows communication with the SerialConnection services on the server side like a local one. This would allow any service on the Internet to remotely handle the serial port traffic, which is very useful for debugging or auditing deployed IoT systems.

(ii) On the NoT Bridge Interpretation layer, a ZigbeeGateway bundle looks for newly created SerialConnection services, checks if they correspond to the Zigbee USB dongle, and creates ZigbeeGateway services implementing the AT command protocol in such a case. Again, the ZigbeeGateway service is defined in the ZigbeeGateway middle interface, which isolates layers and allows using different versions of the USB dongle.

(iii) Above that, there is a ZigbeeDriver, which uses the ZigbeeGateway to accomplish network management tasks, such as network creation or device enumeration, and a Driver Manager which uses the ZigbeeDriver to perform automatic maintenance tasks such as message route maintenance. It would be possible to aggregate these services into a single one but that will introduce complexity and hinder interoperability, whereas using separated services increases reliability and robustness as units of code are smaller and easier to test. This allows also deploying gateway facilities to accomplish transversal tasks, such as network monitoring, debug, maintenance, and commissioning, which may use limited parts of the Driver Layer. In the case of having other networks, the scheme is similar. The ZigbeeDriver will create ZigbeeNode services representing nodes in the network which allows sending and receiving messages to and from the physical device and provides Zigbee related properties such as its MAC address or its type.

(iv) On the CTP Devices layer, a CTP Driver service will use the ZigbeeNode services to create CTP Device services representing the real devices available on the NoT, following the CTP definition. This layer isolates devices' technology and registers them attending to their nature in order to provide interoperability; a temperature sensor always provides temperature the

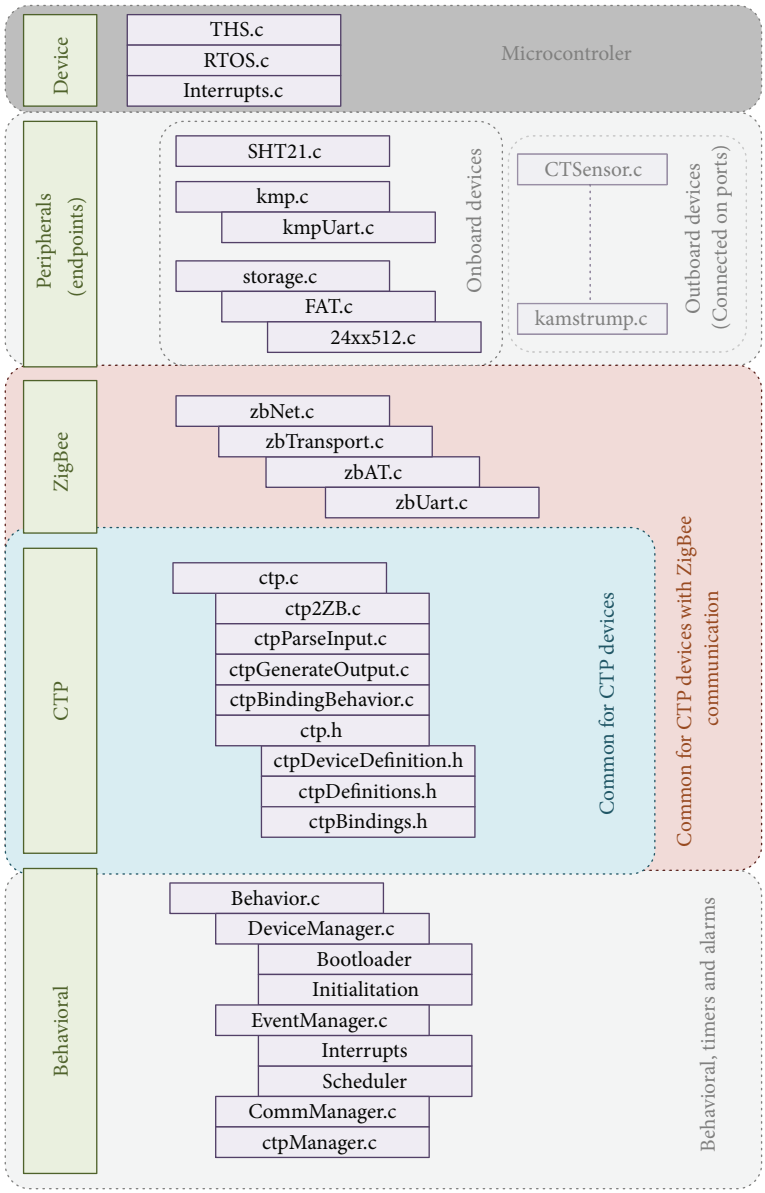


Figure 7: ZigBee *thing* firmware implementation.

same way regardless its technology. There are also different gateway facilities at this level (web services and TCP sockets) providing Internet access and virtual representation in order to allow remote applications to use NoT infrastructure.

(v) At the IoT services layer, a Data Collector service tracks for CTP Device services of type Sensor, subscribes itself to receive a notification when a new sensor value is received, and sends a data report to a remote database, where it can be accessed from elsewhere.

At this point, local network devices, which were not able to reach the IoT by themselves, are now represented by OSGi services which are smart enough to operate on the IoT and among them. Internet applications accessing those *things* require address resolution services which allow reaching them through an URL [2]. This is overcome by delegating *things*' addresses resolution to the IoT gateway which forwards incoming requests to the physical *thing*. Thus, accessing *things* from the Internet is granted by knowing the IoT gateway address. It is also feasible that *things* themselves access the IoT services directly; for example, we feed data from sensors to IoT services on the cloud specialized on data managing like Cosm [43] or Nimbits [44].

On top of the middleware described and thanks to the OSGi modularity, we also have developed communications terminal to sniff ZigBee communication (mainly intended as

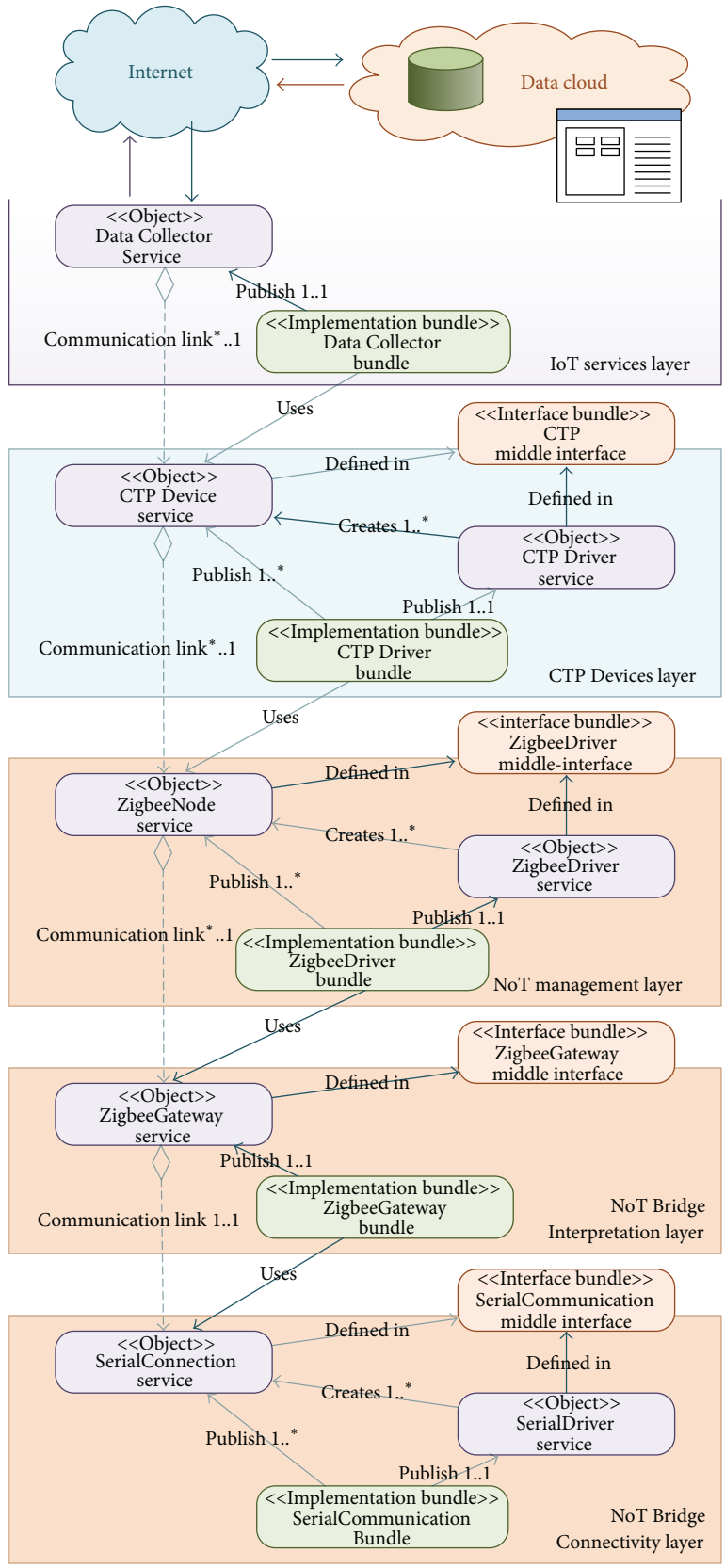


Figure 8: IoT gateway middleware architecture.

a development tool for hardware/firmware/software developers) and a Network Manager that allows full management and commissioning of deployed ZigBee networks.

5.4. System Performance. The system deployed is built by a backbone of 19 nodes (1 coordinator, 17 routers, and 1 data sink) that exchange network management messages. Also 90 sleepy sensors poll their parents every 10 seconds and report the data sink every 10 minutes. Due to network topology, some sensor messages must be relayed by routers up to five times to get to the data sink which passes them to the IoT gateway that maintains network's virtual image, checks data inconsistencies, registers loss of expected messages, and uploads data to exploitation database. These processes allow network commissioning, node-problem targeting, and data integrity validation.

Evaluating the quality of service (QoS) of an IoT application is not easy as it is a large and complex system based on heterogeneous technology and high application dependency. Currently, there are not well-defined standard metrics that can be used as a common agreed and widespread comparison of IoT systems; derived from this application specificity, researchers usually define their own metrics [45–47]. As IoT is usually made up by different subsystems, it is possible to analyze the layers separately; for example, there are many metrics to evaluate the effectiveness of data transfer in networks [48]; however, the success of an IoT should be assessed as a whole not just one particular aspect of the architecture [49].

The system proposed uses 90 sensors to deterministically measure ambient and power consumption information and make it available in the IoT. Three different areas are assessed: relative and absolute energetic cost, data efficiency, and message latency.

Data efficiency considers the amount of data generated in the ZigBee network and the data correctly stored in the cloud database. Similar indicators are used to assess the quality of a communication network, in which case it is common to consider lost messages and total messages. However, for the global evaluation of the IoT is preferable to consider both ends of the system, thus we use the messages generated by the physical *things* and the amount of data available in the logical scope. Note that if there would be nondeterministic events (e.g., alarms, presence detection, etc.) the data generated in the IoT will not be known a priori and also difficult to measure. We define the data efficiency of the IoT, $\mathrm{DE}_{\mathrm{IoT}}$, as

$$\mathrm{DE}_{\mathrm{IoT}} = \frac{\mathrm{DA}_{\mathrm{IoT}}}{\mathrm{DG}_{\mathrm{NoT}}} = \frac{\mathrm{DA}_{\mathrm{IoT}}}{\sum_{i=1}^{n} \left(\mathrm{MS}_i \times \mathrm{ND}_i\right)}, \tag{1}$$

where $\mathrm{DA}_{\mathrm{IoT}}$ is the data available on the IoT, $\mathrm{DG}_{\mathrm{NoT}}$ is the data generated by the NoT, n is the number of *things*, MS_i is the number of messages sent by *thing* I, and ND_i is the amount of data sent in each message (we assume this is constant for all the messages sent by a *thing*). This value can be measured accumulated or disaggregated in different periods to assess the variation of stability over the time. In our case, we check the number of new registers in database and considering that the total number of inputs should be 12960 per day (defined by the number of *things* and their scheduling), we can calculate the ratio. In our case, the accumulated data efficiency is 98.3%. It has been found that in general, errors are due to failures on the electrical supply or Ethernet network not directly attributable to the system.

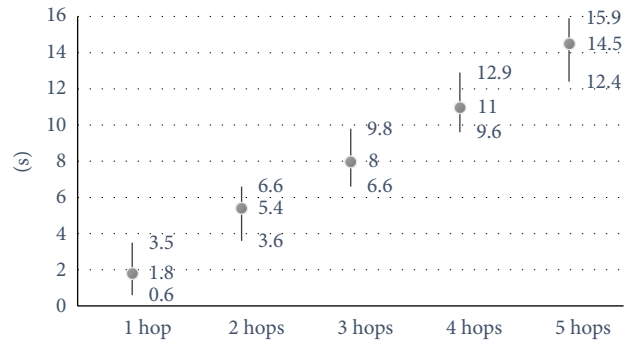


Figure 9: Latency time intervals.

Message latency quantifies the availability of a *thing* in order to exchange information with it. Again, we consider this availability from the IoT layer, that is, how much time takes to force a sensor to refresh, and effectively update the data in the cloud. There are two main factors that influence this indicator: how many hops away from the gateway is the sensor and how much time does the sensor spend in sleep mode (i.e., disconnected from the ZigBee network). Thus, we define the message latency per hop, MLH_j, as:

$$\mathrm{MLH}_j = \frac{\sum_{i=1}^{n_j} \mathrm{RT}_{ij}}{n_j}, \tag{2}$$

where RT_{ij} is the response time for *thing* i, which is j hops away from the gateway, and n_j is the number of *things* being j hops away from the gateway. Figure 9 shows the latency intervals (in seconds) within a 95% confidence interval according to the distance in hops between sensor and gateway. If we need to provide a global metric of the system latency we would say it will be between 0,6 s and 15,9 s.

Energy consumption of a device is a common indicator of its efficiency. In the particular case of IoT, we must take into consideration the consumption of the *things* andalso communications infrastructure (routers) and logical infrastructure (gateway). Given the data volume managed, energy consumption of routers and gateway is independent from the amount of data sent by *things*, and the absolute energetic cost of the system over time $\mathrm{AEC}(t)$ can be defined as follows:

$$\mathrm{AEC}\,(t) = \left[P_{\mathrm{Gtw}} + n_R \times P_R + \sum_{i=1}^{n} P_i\right] \times t, \tag{3}$$

where P_{Gtw} is the power consumption of the gateway, P_R is the power consumption of a router, n_R is the number of routers, P_i is the power consumption of *thing* i, and n is the number of *things*. As we show in [41] to calculate a *thing's* power consumption, a good characterization of its duty cycle is required. We have defined the figure of merit of power

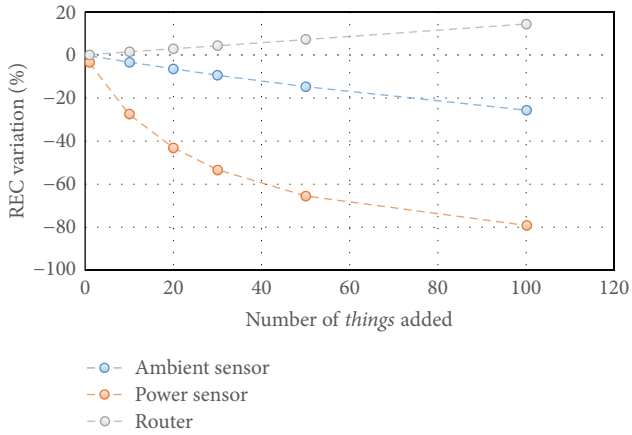


Figure 10: Relative energetic cost variation with respect to the proposed scenario when including new devices.

consumption for each type of *thing* and the routers and we consider constant the power consumption of the gateway. For the proposed IoT, the absolute energetic cost in one day is 19,76 MJ (5,49 kW·h). It is important to remark that only 0,013% of the energy is due to *things*, battery-powered devices. Given that in the proposed scenario, the number of data and time are directly related, we can also define the relative energetic cost REC as the energy required per byte of processed data during a day:

$$\mathrm{REC} = \left.\frac{\mathrm{AEC}(t)}{\mathrm{DA}_{\mathrm{IoT}}}\right|_{1\,\mathrm{day}}. \quad (4)$$

For the present IoT application we have a relative energetic cost of 121,97 J per byte sent. This includes 15 mJ corresponding to the energy drained from batteries. This parameter is related to the scalability of the network; if we enlarge the communications infrastructure (more routers to have broader coverage) and increase the number of *things* (to have more measurement points), relative energetic cost varies as in Figure 10.

Including new sensors supposes decreasing the REC as they provide additional data with reduced energy consumption. Specifically, power sensor consumption is related to a better REC than ambient sensor as it sends more data (44 bytes versus 4 bytes) with a similar energy consumption. As expected, including routers improves network backbone and/or range but worsens the metric, as they do not increase $\mathrm{DA}_{\mathrm{IoT}}$.

6. Discussion

Main conclusion of the described field trial is that architecture and the protocol proposed are feasible in large and long-term IoT deployments. Also it demonstrates that implementation of CTP over ZigBee in order to create low power *things* (running with batteries for a year) that efficiently integrates in an IoT system is possible. As we developed the whole system, from the hardware design of *things* to the service implementation, our concerns have been from the limited resources of hardware and firmware to the versatility and generality required by applications.

In order to analyze under which circumstances CTP is a convenient option, Table 1 compares it with main IoT protocols mentioned in Section 2. Three different levels are identified: (1) standard communication protocols (6lowPAN, RFID, WiFi, Cellular, ZigBee, and Bluetooth), (2) protocols abstracting from communication media and being mounted over the protocol's payload (IEEE1451, CTP), and (3) protocols assuming IP connectivity on any*thing* (CoAP, RESTfull-HTTP, XMPP, MQTT, SensorWeb, EEML, and SenseWeb). Comparison items are as follow.

(i) Application layer interoperability among *things* indicates whether *things* can effectively exchange information among them, for example, if a light controller (protocol A) can be operated by a light sensor (protocol B). Communication standards (1) are limited on this because they are not intended to define how to interoperate with other standards. On the other side, protocols abstracting from communication standard (2, 3) are precisely designed to enable this feature. Also IP-based protocols (3) are more restrictive as they need that *things* are IP enabled.

(ii) *Things*' representation models indicate if the protocol provides a virtual representation of *things*; for example, a temperature sensor has some characteristics (range, accuracy, etc.) and provides some methods (get temperature measurement, configure it, etc.). The protocols that just consider data exchange (e.g., 6LowPAN, WiFi, or MQTT) do not provide this definition hindering interoperability. ZigBee and Bluetooth define some *things*, those considered in their applications. IEEE1451, CTP, and SensorWeb define how any*thing* needs to be specified, for example, the datasheet format.

(iii) *Things*' interaction model: interaction means on step further than representation as it implies providing relationships among *things*, for example, how a light controller can be operated by a light sensor.

(iv) Suitability for low power and lossy networks: this relies very much on the possibility to run on wireless sensor network standards, mainly ZigBee and 6LowPAN.

(v) Simplicity and exhaustiveness are important features that determine adoption of protocols. For example, IEE1451 is a very exhaustive protocol defining everything in a sensor, but precisely this makes its use complicated by an app developer that just wants temperature value of a sensor.

According to Table 1, most similar protocols are CTP, IEEE1451, and ZigBee Cluster Library. In order to make a more detailed comparison among them, we define a scenario based on a ZigBee temperature sensor where these three specifications are encapsulated in the payload of the ZigBee application layer (Figure 11).

Table 1: IoT protocols comparison.

		Application layer interoperability among things	*Things'* representation models	*Things'* interaction model	LLN (Low Power and Lossy Networks) suitability	Simplicity	Exhaustiveness and level of detail
1	6lowPAN, RFID, WiFi, and Celular	Limited to standard (stack layer)	No	No	6lowPAN: excellent Others: fair	Not applicable (App. layer not defined)	Not applicable (App. layer not defined)
	ZigBee Cluster Library, Bluetooth Profiles	Limited to standard (App. layer)	Yes, but limited to specific applications	Yes, but limited to specific applications	ZigBee: excellent Bluetooth: fair	Good	Excellent
2	IEEE1451	Full (delegated to NoT protocol)	Yes, but limited to sensor and actuators	Yes, but limited to sensor and actuators	Excellent	Good	Excellent
	CTP	Full (delegated to NoT protocol)	Yes	Yes	Excellent	Excellent	Good
3	CoAP, RESTfullHTTP, XMPP, MQTT	Full (delegated to IP protocol)	No	No	CoAP: good (UDP) Others: fair (TCP)	Not applicable (App. layer not defined)	Not applicable (App. layer not defined)
	SensorWeb, EEML, SenseWeb	Full (delegated to IP protocol)	Yes	No	Fair (TCP)	EEML: Excellent Others: Fair	EEML: Good Others: Excellent

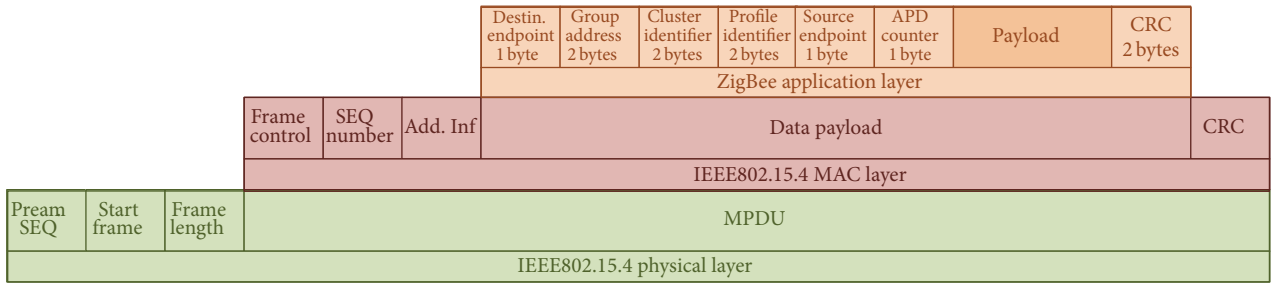


Figure 11: ZigBee protocol encapsulation.

Communication is established through requests and corresponding responses; Table 2 contains the frame descriptions for the three cases. Each frame includes different fields defined in the specifications.

In order to use the temperature sensor several requests need to be answered by the sensor: (1) retrieving general information about the device, (2) retrieving channel/endpoint mandatory information, and (3) retrieving temperature measurement. If using ZCL, sensor temperature must be 2 bytes in length, in degrees Celsius, and specified in the range from −273.15°C to 327.67°C with aresolution of 0.01°C. Thus, step 2 is not needed as information about the measure is already defined. This is positive in terms of easy use but limitative in some cases; if using CTP or IEEE 1451, variables such as accuracy, range, units, and so forth are open to the hardware developer and must be specified in the channel/endpoint mandatory information. Table 3 indicates the number of bytes exchanged for each action and the number of variables encoded. It should be noted that just mandatory fields are considered in the table.

Regarding protocol efficiency, it is obvious that ZCL is the most efficient, but this has important implications: as ZCL focuses on specific application domains (home and building automation, health, etc.) if the *thing* is not specified in the standard or the representation of the information differs from it, its implementation is not possible and thus it will not be interoperable. For example, there is no way to use inertial sensors according to ZCL.

On the other hand, IEEE 1451 is the heaviest protocol because it leaves open a lot of variables. This flexibility turns to be its main drawback to be used in IoT applications because it makes it too complicated for developers not versed into electronics. Additionally, as it is more oriented to electronics than to application, it does not consider human-machine interfaces in its specification, just sensors and actuators.

Table 2: Frame description of IEEE 1451, CTP, and ZCL.

		Header	Payload
IEEE 1451	Request (6 + *n*) bytes	Destination transducer channel: 2 bytes Command Class: 1 byte Command Function: 1 byte Offset Length: 2 bytes	Data: *n* bytes
	Response (3 + *n*) bytes	Success flag: 1 byte Offset Length: 2 bytes	Data: *n* bytes
CTP	CTP Frame (5 + *n*) bytes	Endpoint destination: 1 byte Length: 2 bytes Cluster identification: 1 byte Command identifier: 1 byte	Data: *n* bytes
ZigBee Cluster Library	ZCL Frame (5 + *n*) bytes	Frame Control: 1 byte Manufacturer code: 2 bytes Transaction sequence number: 1 byte Command identifier: 1 byte	Id attribute: 2 bytes Value attribute: *n* bytes ⋮ Id attribute *m*: 2 bytes Value attribute *m*: *n* bytes

Table 3: Number of payload bytes (differentiating request + response) and variables associated.

	IEEE 1451		CTP		ZigBee Cluster Library		Observations
	Bytes	Variables	Bytes	Variables	Bytes	Variables	
Retrieve general information about the device	50 (7 + 43) 42 (7 + 35)	21	45 (5 + 31)	5	14 (7 + 7)	1	In case of ZCL, information obtained is partial, same amount/type of variables must be done from a lower level of the protocol; at ZigBee Device Object.
Retrieve channel/endpoint mandatory information	111 (7 + 104) 33 (7 + 26)	18	34 (5 + 29)	7	—	—	ZCL strictly defines the nature of the variables in each cluster, so no need to declaration.
Retrieve temperature measurement	12 (7 + 5)	1	12 (5 + 7)	1	14 (7 + 7)	1	

As a summary, CTP is a balanced protocol especially suited for the IoT allowing definition of any device and just specifying those variables that are needed at application level. As it can be efficiently encapsulated in any communication protocol, it is strongly focused on the interoperability of *things* while considering their technical limitations (energy, processing, and communication capabilities). CTP is based on an ontological representation and interaction model; thus, besides interoperability, it allows local intelligence and interaction among *things*. Also as it describes the characteristics and properties of *things*, higher-level protocols can use it as a source of information.

7. Conclusions

IoT is perceived as a huge generator of services and applications, but it requires that two important issues to be solved: *things*' connectivity (communication of the physical *things* with the Internet) and interoperability at all levels (understanding of the information exchanged). This paper emphasizes the need to maintain an integrative and broad point of view when considering architectures and protocols for IoT. Considering IoT development areas and their associated technologies, an architecture and protocol with a comprehensive view considering current technological capabilities at all levels (services, gateway communications, firmware, and hardware) are proposed.

In order to tackle connectivity challenge, the appropriateness of using gateway-based solutions to connect Networks of *Things* with the Internet has been discussed. This is considered an optimal way to solve the problem of the heterogeneous networks, jumping from the real world to the IP world and ensuring the unambiguous designation of each of the *things* in the Internet, performing a tunneling process (e.g., the need of a unique identification). Regarding the functionality of the IoT gateway, the use of layered and modular middleware architecture to constitute a versatile, interoperable, scalable, and easy to maintain system has been proposed.

Common *Things* Protocol (CTP) has been presented as a solution to provide interoperability among *things*. Main guidelines considered in its design are an ontological representation and interaction model of *things* andimplementation feasibility in standard communication protocols. CTP reflects the need for equilibrium between the networks of *things*, data traffic, and virtualized world of *things*, which

is not common in the current proposals. Main principles of CTP are self-description of the nature and capabilities of the *thing*, organization capabilities through the endpoint and cluster concepts, simplifying the use of the *thing* in standard situations through mechanisms such as modes and scenarios, allowing distributed intelligence by using bindings and behaviors, and being simple and compact to ease its implementation fitting on current standard communication protocols.

The IoT architecture (ZigBee network of *things*, IoT gateway and related Internet services) and CTP proposal have been implemented in two European research projects related to smart environments for Ambient Assisted Living and energy efficiency. The paper describes specific hardware and software implementations and deployment in large scale environments, with real end users, during periods of several months. We also measure energetic cost, data efficiency, and message latency metrics demonstrating proposal's feasibility and suitability in order to meet current IoT requirements.

This paper does not intend to define a standard but an initial proposal to adopt a minimum agreement (currently not considered in the state of the art) with a global and inclusive vision that facilitates the development of the IoT. To this end, it should be mentioned that most of the developments shown have been carried out under open source license, in particular CTP is accessible and well documented [50].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] I. T. Union, "ITU Internet Reports 2005: The Internet of *Things*," 2005.

[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of *Things*: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[3] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton, "Smart objects as building blocks for the internet of *things*," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2010.

[4] R. V. Kulkarni, A. Förster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 1, pp. 68–96, 2011.

[5] ISTAG, "Orientations for EU ICT R&D & Innovation beyond 2013-10 keys recomendation," July 2011, http://cordis.europa.eu/fp7/ict/istag/documents/istag_key_recommendations_beyond_2013_full.pdf.

[6] M. Swan, "Sensor Mania! The Internet of *Things*, Wearable Computing, Objective Metrics, and the Quantified Self 2. 0," *Journal of Sensor and Actuator Networks*, vol. 1, no. 3, pp. 217–253, 2012.

[7] "sensei (Real World Dimension of the NEtwork of the Future)," http://www.ict-sensei.org/index.php.

[8] "IoT- A, (Internet of *Things* Architecture)," http://www.iot-a.eu/public.

[9] "Internet of *Things* Strategic Research Roadmap," European Commission-Information Society and Media DG-BU31 01/18 B-1049, Brussels, Belgium, 2009.

[10] D. Uckelmann, M. Harrison, and M. Florian, "An architectural approach towards the future internet of *things*," in *Architecting the Internet of Things*, pp. 1–24, Springer, 2011.

[11] D. Miorandi, S. Sicari, F. de Pellegrini, and I. Chlamtac, "Internet of *things*: vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[12] MQTT, http://mqtt.org/.

[13] "ipSo Alliance," http://www.ipso-alliance.org/.

[14] T. Usländer, A. Berre, C. Granell et al., "The future internet enablement of the environment information space," in *Proceedings of the International Symposium on Environmental Software Systems (ISESS '13)*, Neusiedl am See, Austria, 2013.

[15] C. Gómez, J. Paradells, and J. Caballero, *Sensors Everywhere. Wireless Network Technologies and Solutions*, Fundación Vodafone España, 2010.

[16] A. Kapadia, S. Myers, X. Wang, and G. Fox, "Secure cloud computing with brokered trusted sensor networks," in *Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS '10)*, pp. 581–592, May 2010.

[17] A. Ibarz, G. Bauer, R. Casas, A. Marco, and P. Lukowicz, "Design and evaluation of a sound based water flow measurement system," *Lecture Notes in Computer Science*, vol. 5279, pp. 41–54, 2008.

[18] ZigBee Specification, "ZigBee Cluster Library Specification," (053474r17), 2008.

[19] Y.-F. Lee, H.-S. Liu, M.-S. Wei, and C.-H. Peng, "A flexible binding mechanism for ZigBee sensors," in *Proceedings of the 5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP '09)*, pp. 273–278, December 2009.

[20] C. Bormann, A. P. Castellani, and Z. Shelby, "CoAP: an application protocol for billions of tiny internet nodes," *IEEE Internet Computing*, vol. 16, no. 2, pp. 62–67, 2012.

[21] XMPP Standards Foundation, "XMPP Standard," http://xmpp.org/.

[22] "Constrained RESTful Environments (core)," https://datatracker.ietf.org/wg/core/charter/.

[23] "How Can IEEE, 1451 Be Applied," June 2007, http://ieee1451.nist.gov/.

[24] I. I. a. M. Society, *IEEE Standard for a Smart Transducer Interface for Sensors and Actuators*, 2007.

[25] D. Sweetser, V. Sweetser, and J. Nemeth-Johannes, "A modular approach to IEEE-1451.5 wireless sensor development," in *Proceedings of the IEEE Sensors Applications Symposium*, pp. 82–87, February 2006.

[26] J. Higuera, J. Polo, and M. Gasulla, "A ZigBee Wireless sensor network compliant with the IEEE 1451 standard," in *Proceedings of the IEEE Sensors Applications Symposium (SAS '09)*, pp. 309–313, February 2009.

[27] J. E. Higuera and J. Polo, "IEEE 1451 standard in 6LoWPAN sensor networks using a compact physical-layer transducer electronic datasheet," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2751–2758, 2011.

[28] Z. Wemlinger and L. Holder, "The COSE ontology: bringing the semantic web to smart environments," in *Toward Useful Services for Elderly and People with Disabilities*, vol. 6719, pp. 205–209, Springer, 2011.

[29] D. Bonino, E. Castellina, and F. Corno, "DOG: an ontology-powered OSGi domotic gateway," in *Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '08)*, pp. 157–160, November 2008.

[30] W. Wang, S. De, G. Cassar, and K. Moessner, "Knowledge representation in the internet of things: semantic modelling and its applications," *Automatika*, vol. 54, no. 4, pp. 388–400, 2013.

[31] B. Christophe, "Semantic profiles to model the 'web of *things*'," in *Proceedings of the 7th International Conference on Semantics, Knowledge, and Grids (SKG '11)*, pp. 51–58, October 2011.

[32] E. Y. Song and K. Lee, "Understanding IEEE 1451—networked smart transducer interface standard—what is a smart transducer?" *IEEE Instrumentation and Measurement Magazine*, vol. 11, no. 2, pp. 11–17, 2008.

[33] R. Johnson and S. P. Woods, "Proposed enhancements to the IEEE, 1451. 2 standard for smart transducers," September 2009, http://archives.sensorsmag.com/articles/0901/74/main.shtml.

[34] M. Compton, P. Barnaghi, L. Bermudez et al., "The SSN ontology of the W3C semantic sensor network incubator group," *Web Semantics*, vol. 17, pp. 25–32, 2012.

[35] L. Feng, P. M. G. Apers, and W. Jonker, "Towards context-aware data management for ambient intelligence," in *Database and Expert Systems Applications*, vol. 3180 of *Lecture Notes in Computer Science*, pp. 422–431, Springer, 2004.

[36] S. Petersen and S. Carlsen, "WirelessHART versus ISA100.11a: the format war hits the factory floor," *IEEE Industrial Electronics Magazine*, vol. 5, no. 4, pp. 23–34, 2011.

[37] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[38] European 6th Framework Programme for Research and Technological Development, "Easy Line Plus," http://www.easylineplus.com/.

[39] C. P. E. Commision, "Renaissance," http://www.renaissance-project.eu/?lang=en.

[40] U. o. Zaragoza, "Liga Energetica edificio I+D+i," http://proyectosostenibilidad.unizar.es/institutos/index.php/monitorizacion.

[41] Á. Asensio, R. Blasco, Á. Marco, and R. Casas, "Hardware architecture design for WSN runtime extension," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 136745, 11 pages, 2013.

[42] O. A. website, "OSGi Alliance website," http://www.osgi.org/Main/HomePage.

[43] http://www.cosm.com.

[44] http://www.nimbits.com.

[45] M. W. Ryu, J. Kim, S. S. Lee, and M. H. Song, "Survey on internet of *things*: toward case study," *Smart Computing Review*, vol. 2, no. 3, 2012.

[46] M. Paschou, E. Sakkopoulos, E. Sourla, and A. Tsakalidis, "Health Internet of *Things*: metrics and methods for efficient data transfer," *Simulation Modelling Practice and Theory*, vol. 34, pp. 186–199, 2013.

[47] A. P. Castellani, N. Bui, P. Casari, M. Rossi, Z. Shelby, and M. Zorzi, "Architecture and protocols for the internet of *things*: a case study," in *Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM '10)*, pp. 678–683, April 2010.

[48] S. Forsström, P. Österberg, and T. Kanter, "Evaluating ubiquitous sensor information sharing on the internet of *things*," in *Proceedings of the IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, 2012.

[49] J. Gubbia, R. Buyyab, S. Marusic, and M. Palaniswami, "Internet of *Things* (IoT): a vision, architectural elements, and future directions," *Future Generation Computer Systems*, no. 29, pp. 1645–1660, 2013.

[50] HOWLab, 2013, http://openlab.unizar.es/index.php/conocimiento/ctp.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 576573, 14 pages
http://dx.doi.org/10.1155/2014/576573

Research Article

Energy-Efficient Node Selection Algorithms with Correlation Optimization in Wireless Sensor Networks

Hongju Cheng,[1] Zhihuang Su,[1] Daqiang Zhang,[2] Jaime Lloret,[3] and Zhiyong Yu[1]

[1] *College of Mathematics and Computer Science, Fuzhou University, Fuzhou, Fujian 350108, China*
[2] *School of Software Engineering, Tongji University, Shanghai 201804, China*
[3] *Department of Communications, Polytechnic University of Valencia, Valencia, Camino de Vera 46022, Spain*

Correspondence should be addressed to Hongju Cheng; cscheng@fzu.edu.cn

Received 21 December 2013; Accepted 27 January 2014; Published 27 March 2014

Academic Editor: Joel J. P. C. Rodrigues

The sensing data of nodes is generally correlated in dense wireless sensor networks, and the active node selection problem aims at selecting a minimum number of nodes to provide required data services within error threshold so as to efficiently extend the network lifetime. In this paper, we firstly propose a new Cover Sets Balance (CSB) algorithm to choose a set of active nodes with the partially ordered tuple (data coverage range, residual energy). Then, we introduce a new Correlated Node Set Computing (CNSC) algorithm to find the correlated node set for a given node. Finally, we propose a High Residual Energy First (HREF) node selection algorithm to further reduce the number of active nodes. Extensive experiments demonstrate that HREF significantly reduces the number of active nodes, and CSB and HREF effectively increase the lifetime of wireless sensor networks compared with related works.

1. Introduction

A wireless sensor network consists of spatially sensor nodes which are generally self-organized and connected by wireless communications [1]. Today such networks are used in many industrial and consumer applications, such as traffic data collection, vehicular monitoring and control, security surveillance, and smart homes. Each sensor node is equipped with a sensing device which can detect the environmental condition. The nodes are also powered by limited batteries and it is difficult or impossible to replace them in some special environments. It is why energy efficiency is always the most important criterion for such networks. One important approach to extend the network lifetime is to reduce the number of required packet transmissions in the network [2–5], such as clustering [6–11], in-network data aggregation [12–18], and approximate data collection [19, 20]. In these scenarios, all nodes in the network are considered active and the data are gathered from all nodes during the collecting process.

However, it is not an efficient way to collect all raw data from each node in some special applications which aim to collect information originated from the environment, such as temperature, humidity, and pressure. In these applications, it is fully tolerant if the final collected information is just within error threshold. The sensing data of each node is generally a noise version of the observed phenomenon and there is a deviation among them due to distance, location, or node sensitivity. Nodes are generally correlated if they are observing the same physical phenomena. Correlations between nodes are described in some simple ways such as the maximum or minimum value between nodes [21]. In this paper, correlation occurs if the sensing data of simple node can be obtained from the other nodes. Accordingly, a subset of active nodes can be selected to provide the required sensing service within error threshold, and the rest nodes can go to sleep and preserve energy. In this way, the active node selection strategy with correlation optimization not only prolongs the network lifetime, but also helps to solve other issues in dense wireless sensor networks [22], such as lower network throughput, serious node conflict, and excessive packet transmissions.

How to describe the correlation among the sensing data quantitatively is the key issue when achieving an efficient

active node selection strategy. The distance function is generally considered as an important model to formulate the data similarity between nodes because the sensitivity is sometime related to the distance between the source and sensing device. Here we adopt Manhattan distance between sensing data as error metric [22]. Based on the observation that the sensing data are similar to each other if they are close enough, Kotidis [23] proposes Snapshot query in which only selected active nodes report their sensing data, and sensing data of one-hop nodes is computed by active nodes. Liu et al. [24] propose an EEDC algorithm which divides the nodes into disjoint cliques based on spatial correlation so that the nodes in the same clique have similar sensing data and can communicate directly with each other. Hung et al. [22] propose a DCglobal algorithm to determine a set of active nodes with high energy levels and wide data coverage ranges.

Figures 1(a) and 1(b) show the selected nodes with EEDC and DCglobal for a given wireless sensor network, where each circle denotes one node (the sensing data value is marked above the circle). The edge between a pair of nodes denotes that they can communicate directly with each other. Here we assume that Manhattan distance is used as the similarity function and the error threshold is 0.5. The selected active nodes are marked with black solid circle. The selected active node set with EEDC is $\{s_1, s_2, s_5, s_6, s_8, s_9, s_{12}\}$, and it is $\{s_1, s_5, s_{10}, s_{13}\}$ with DCglobal. According to Figure 1, the number of selected nodes with DCglobal is smaller than EEDC in this example.

The concept of data coverage range is firstly introduced to describe the correlation among nodes and defined as a node set in which the distance between each element and the given node is within the error threshold [22]. In fact, it is a simple extension of one-hop data coverage [23]. Another issue of [22] is the efficiency of proposed node selection algorithm. The partially ordered tuple (residual energy, data coverage range) is used to select an active node set, which ensures that the selected nodes always have high reserved energy, but the number of selected active nodes is not minimized.

To address these problems, we introduce several new concepts, that is, cover set, active node, and covered node, and propose a new Cover Sets Balance algorithm (CSB) to choose a set of active nodes with wide data coverage range and high energy level by using the partially ordered tuple (data coverage range, residual energy) and build the corresponding cover set in sequence to ensure the selected active nodes have high residual energy. In this way, the set of final selection nodes generally owns larger residual energy and smaller size, which helps to extend the network lifetime. Figure 1(b) demonstrates the set $\{s_1, s_5, s_{10}, s_{13}\}$ generated by DCglobal assuming that reserved energy is identical to all nodes in the network, which is similar to the partially ordered tuple (data coverage range, residual energy). Figure 1(c) demonstrates the result as $\{s_3, s_5, s_{10}, s_{13}\}$ with the proposed CSB algorithm in this paper. $\{s_1, s_7, s_8\}$ is a cover set for node s_3 and each node in the set is a feasible candidate regarding s_3.

In the following we show some nodes can be further removed from the selected active node set with CSB. As shown in Figure 1(c), the sensing data of s_3, s_5, s_{10} is 35.5, 36.1, and 34.5, respectively, the average value of s_5 and s_{10} is 35.3. The Manhattan distances between 35.3 and sensing data s_1, s_3, s_7, and s_8 are 0.1, 0.2, 0.1, and 0.3 accordingly (they are all less than the error threshold 0.5). It means that the sensing data of $CS_3 + \{s_3\}$ is computed by s_5 and s_{10}. Accordingly, s_3 is removed and then we have a smaller active node set $\{s_5, s_{10}, s_{13}\}$, as shown in Figure 1(d). Following this observation, we introduce a novel concept Correlated Node Set (CNS) and then we propose a High Residual Energy First (HREF) node selection algorithm to reduce the number of active nodes. The main contributions of this paper are as follows.

(i) We propose a Cover Sets Balance algorithm (CSB) to select a set of active nodes with wide data coverage ranges and high energy levels. In each active node selection step, we use the partially ordered tuple (data coverage range, residual energy) to find an initial active node set and then balance the size of the cover sets in order to replace low-energy nodes.

(ii) We propose a Correlated Node Set Computing algorithm (CNSC) to calculate the correlated node set with minimum set size and maximum geometric mean of residual energy of each node in the sensor network by following the observation that some nodes selected by CSB can be further removed.

(iii) We propose a High Residual Energy First algorithm (HREF) to reduce the number of active nodes selected with CSB by removing nodes which can be computed by correlated node sets.

The rest of this paper is organized as follows. In Section 2, we describe the system model. Section 3 introduces CSB and HREF algorithms. The theoretical analysis of the algorithms is proposed in Section 4. In Section 5, we describe the simulation results and performance analysis. Section 6 presents the related works and Section 7 is conclusion.

2. System Model

A wireless sensor network generally consists of a set of stationary nodes $V = \{s_1, s_2, \ldots, s_n\}$, and each node in the network has identical transmission radius r. The network is formulated as an undirected graph $G = (V, E)$ with V as the set of nodes and E as the set of links. Without loss of generality, both s_i and i are used to represent one single node in the network. There is a link (i, j) between node i and j if they communicate with each other directly.

The nodes are equipped with unreplaceable or unrechargeable batteries. The reserved energy for node i at time t is denoted by $e_i(t)$. The collected data from one single node is a noise version of the practical phenomena. In these applications, the collected information from the sensor network is tolerant in case that it is within a given error threshold ε.

The notations used in this work are listed as the following:

V: Set of nodes in the network

n: Number of nodes in the network

ε: One given error threshold

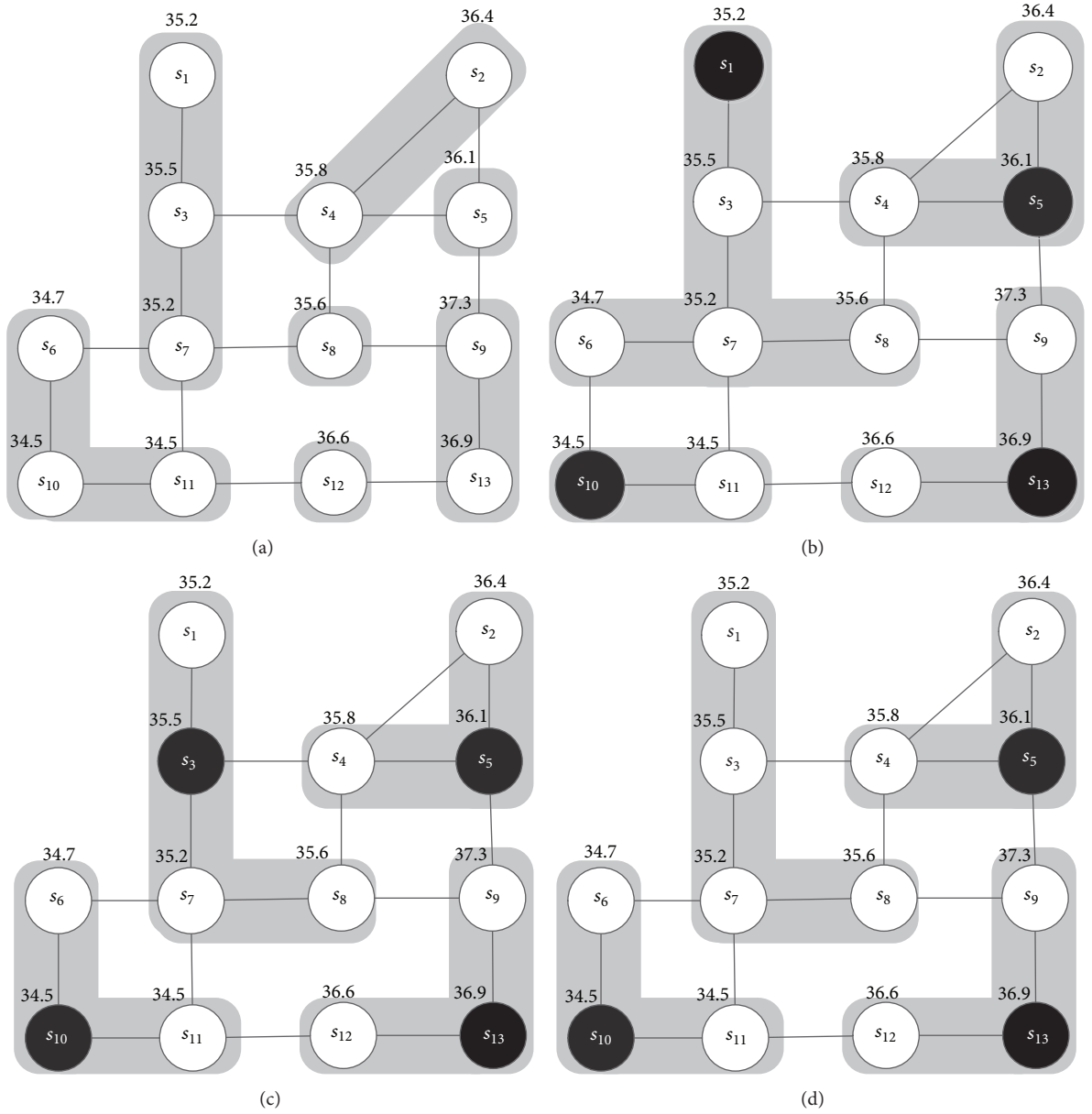


Figure 1: An example to demonstrate different algorithms. (a) EEDC, (b) DCglobal, (c) CSB, and (d) HREF.

$x_i(t)$: Sensing data of i at time t

$d(x_i(t), x_j(t))$: Distance between the sensing data of i and j

$\text{Energy}_i(t)$: Residual energy of i at time t

DCR_i: Data coverage range of i

CS_i: Cover set of i

CNS_i: Correlated node set of i

ANS: Active node set

r: Transmission radius

$maxd$: Maximal number of nodes in a correlated node set

$\text{Event}_j(t)$: Value of event j at time t

Interval: Interval to reselect a new active node set.

The correlation among sensing data especially in a dense wireless sensor network is helpful to extend the network lifetime. Some researchers studied the correlation between nodes and provided some models [25]. Among all these models, it is common to adopt distance function, such as Manhattan distance $d(x_i(t), x_j(t))$ to represent the correlation between sensing data $x_i(t)$ and $x_j(t)$ at time t [22], which is represented as $d(x_i(t), x_j(t)) = |x_i(t) - x_j(t)|$. Without loss of generality, we follow this correlation model in this paper. Note that our algorithms are adapted to any other correlation models with minor modification.

The sensing data of s_j is called to be *computed* with the sensing data of s_i if the sensing data has a high correlation level; that is, $d(x_i(t), x_j(t)) \leq \varepsilon$, where ε is the given error threshold. S_j is also in the *data coverage range* of s_i. The definitions are given as follows.

Definition 1 (data coverage range (DCR)). Given an error threshold ε in the sensor network, the data coverage range DCR_i of i is a subset of V, in which Manhattan distance from each node to i is no more than ε, and $i \notin \mathrm{DCR}_i$.

For the example in Figure 1(b), $\varepsilon = 0.5$, $\mathrm{DCR}_1 = \{s_3, s_6, s_7, s_8\}$, $\mathrm{DCR}_3 = \{s_1, s_4, s_7, s_8\}$, and so on.

Definition 2 (active node set (ANS) and active node). Given a sensor network $G = (V, E)$, an active node set (ANS) is a subset of V, in which each $i \in V$ either belongs to ANS or one data coverage range DCR_j, where $j \in \mathrm{ANS}$. Any node in ANS is named as an *active node*.

For the example in Figure 1(c), $\mathrm{ANS} = \{s_3, s_5, s_{10}, s_{13}\}$ and each node in the set is an active node.

Definition 3 (cover set (CS) and covered node). Given a sensor network $G = (V, E)$ and according ANS, the cover set CS_i for any given $i \in \mathrm{ANS}$ is a subset of DCR_i, and $\mathrm{CS}_i \cap \mathrm{CS}_j = \emptyset$ in case $i \neq j$. Any node in CS_i is named as a *covered node*.

For the example in Figure 1(c), $\mathrm{CS}_3 = \{s_1, s_7, s_8\}$, s_3 is the active node, and s_1, s_7, and s_8 are covered nodes.

Sensor data is affected by the events in monitored region, and the influence of each event on a sensor is inversely proportional to their distance. Here we assume that correlation occurs among all active nodes in the sensor network.

Definition 4 (correlated node set (CNS)). Given a sensor network $G = (V, E)$ and its corresponding ANS, the correlated node set CNS_i for $i \in \mathrm{ANS}$ is a subset of ANS, and the arithmetic mean $\bar{s}$ for sensing data of nodes in CNS_i satisfies the error threshold condition; that is, $d(x_j(t), \bar{s}) \leq \varepsilon$, $j \in \mathrm{CS}_i + \{i\}$. The sensing data of $\mathrm{CS}_i + \{i\}$ is said to be *computed* by CNS_i.

For the example in Figure 1(d), $\mathrm{CNS}_3 = \{s_5, s_{10}\}$, $\bar{s} = 35.3$, $d(x_1, \bar{s}) = 0.2 \leq \varepsilon$, $d(x_7, \bar{s}) = 0.1 \leq \varepsilon$, $d(x_8, \bar{s}) = 0.3 \leq \varepsilon$, $d(x_3, \bar{s}) = 0.2 \leq \varepsilon$.

Definition 5 (CNS computing problem). Given a sensor network $G = (V, E)$, ANS, sensing data $X = \{x_1, x_2, \ldots, x_n\}$, cover sets $\mathrm{CS} = (\mathrm{CS}_1, \mathrm{CS}_2, \ldots, \mathrm{CS}_n)$, and reserved energy $\mathrm{Energy} = \{e_1, e_2, \ldots, e_n\}$, the CNS computing problem is to find a correlated node set CNS_i for node $i \in \mathrm{ANS}$, and the size of $|\mathrm{CNS}_i|$ is minimized while the geometric mean of residual energy $\hat{e}$ is maximized, where $\hat{e} = \sqrt[n]{e_1 e_2 \cdots e_n}$.

Note that we adopt the geometric average of the residual energy in the correlated node set by following the observation that the average geometric averaging gives higher results for lower variations in the data values for a given data set with a fixed arithmetic [26].

Definition 6 (active node selection problem). Given a sensor network $G = (V, E)$, the sensing data $X = \{x_1(t), x_2(t), \ldots, x_n(t)\}$, reserved energy levels $\mathrm{Energy}(t) = \{e_1(t), e_2(t), \ldots, e_n(t)\}$ at time t and given threshold ε, the active node selection problem is to select a set of active nodes $\mathrm{ANS}(t)$ at time t, where all sensing data in the network can be computed by their corresponding active nodes, and the network lifetime is maximized, that is, $\max\{t\}$.

The active node selection problem is to find the active node set during each epoch and aim at maximizing the network lifetime. The problem is proven to be NP-hard by mapping it to the set covering problem or minimum dominating set problem [26–28]. In this paper, we design two heuristic algorithms, namely, CSB and HREF for this problem.

3. Heuristic Algorithms

3.1. CSB Algorithm. Most related works use the concept of data coverage range combined with energy to solve the active node selection problem. In this section we illustrate the Cover Sets Balance algorithm (CSB) based on the idea of data coverage range.

In data collection process, only active nodes are required to provide perception service, and the rest nodes are closed to preserve energy. An intuitive approach for the node selection process is to use the partially ordered tuple (data coverage range, residual energy) [23]. Another approach is to use partially ordered tuple (residual energy, data coverage range) to select active nodes with higher residual energy [22]. However, the number of selected nodes is generally larger than the former approach, which means that more energy consumption is necessary when providing perception service during the given epoch. Obviously, we need a balance between the two metrics, that is, the data coverage range and residual energy.

The basic idea of Cover Sets Balance (CSB) algorithm is described as the following: (1) generate an initial active node set and the corresponding cover sets through the previous data coverage range priority strategy; (2) replace active nodes with high-energy candidates. Note that the candidates must cover all nodes within the same cover set. For example, in Figure 1(b), $\mathrm{CS}_1 = \{s_3, s_6, s_7, s_8\}$, $\mathrm{CS}_5 = \{s_2, s_4\}$, $\mathrm{CS}_{10} = \{s_{11}\}$, and $\mathrm{CS}_{13} = \{s_9, s_{12}\}$. It is seen that s_1 covers four nodes, namely, s_3, s_6, s_7, and s_8, and thus the candidate node for s_1 must cover the above four nodes too. However, if the cover set is too large, it is possible to find no candidate nodes. The case is similar when the cover set is too small. Obviously we need a new method to provide more candidate nodes so that the network lifetime is extended in an efficient way.

We adopt a cover set balance strategy to balance the set size by moving nodes from larger cover sets to smaller ones. The initial cover sets are sequenced in descending order of the set size, and then we check nodes in one cover set and try to move them to another with smaller size. This

process continues until all sets are checked and finally they are balanced. This strategy is helpful to increase the number of candidate nodes with higher residual energy by cutting down the maximal deviation of each cover set in the balance progress.

The final step of the CSB algorithm is to replace the selected active nodes with candidates by order of reserved energy. In this way, we finally build an active node set with the same size as its initial version but higher residual energy, which is helpful to extend the network lifetime.

The CSB algorithm can be divided into three processes and pseudocodes are shown in Algorithm 1.

The Initialization_Process is used to build a primary active node set and corresponding cover sets. The basic steps are described as follows. There are two different states for each node in the network, namely, *Primary-Covered* and *Un-Covered*, which are used to mark whether it is within the cover set of one node in the active node set. The states for all nodes are initialized as *Un-Covered* (Line 3). Then we sort nodes with partially ordered tuple (data coverage range, residual energy) and initialize the active node set as empty set (Line 4-5). Finally, we check nodes in sequence with state as *Un-Covered*, and add them into the active node set if the required conditions are satisfied (Line 6–11).

The Cover_Set_Balance_Process aims at balancing the size of cover sets generated with the Initialization_Process. Firstly, the cover sets are ordered and checked accordingly to their set size (Line 13). Secondarily, nodes in a given cover set CS_i are sorted into a sequence with descending order of their deviation to i (Line 15), and they are moved to another cover set with smaller size (Line 16–19). This process continues until all nodes in the cover set are checked (Line 14–20).

The Node_Replace_Process focuses on nodes exchange by replacing the low-energy active nodes with high-residual-energy candidates. All feasible candidate nodes of i are checked (Line 23–25), and we select the one (marked as m) with maximal residual energy among all these candidates (Line 26). Finally, the active node set is updated as well as the cover set for node m (Line 27).

The CSB algorithm follows the idea of replacing the active nodes with candidates with higher residual energy. However, it has the same number of active nodes compared with the approach which only uses the partially ordered tuple (data coverage range, residual energy). In the following we introduce a new HREF algorithm to further reduce the number of active nodes based on CNSC algorithm.

3.2. HREF Algorithm. We first introduce an algorithm for the CNS computing problem and then propose a High Residual Energy First node selection algorithm (HREF) for the active node selection problem.

3.2.1. CNSC Algorithm. The CNS computing problem is to find one subset CNS_i for $i \in$ ANS and aim at minimizing CNS_i as well as maximizing the geometric mean of residual energy $\widehat{e} = \sqrt[n]{e_1 e_2 \cdots e_n}$. To find out the optimal CNS_i, an intuitive way is to calculate the average value of sensing data for each subset ANS stored in a sequenced list L. Then, pick out the average values whose deviation is no more than ε and the corresponding correlated node set in the list. Finally, select the CNS_i with minimized node set and maximum geometric mean of residual energy as the final correlated node set for i. Obviously, the above solution is to find the optimal result but has exponential time complexity ($O(2^{|ANS|})$).

To reduce the time complexity, we assume each CNS_i has at most *maxd* nodes, where *maxd* is a given value depending on the network environment. The CNS computing problem is then converted to the problem of selecting at most *maxd* number of nodes in ANS within the error threshold. Then, calculate each subset combined with the selected *maxd* nodes and add its average value into L with the following iteration process: in the ith iteration, the average value for each subset of $\{x_1, x_2, \ldots, x_i\}$ is calculated based on the average value of subset of $\{x_1, x_2, \ldots, x_{i-1}\}$. There are two basic operations in the iteration process, namely, $(L + x)$ and merge $L[L, L + x]$. $(L + x)$ represents the new list by adding x into each element in the initial sequence L, as shown in Form. (1); and merge $L[L, L + x]$ represents the ordered list for the combined result of L and $(L + x)$:

$$L + x = \left\{ \frac{L(i) \times L\_\text{count}(i) + x}{L\_\text{count}(i) + 1} : i \in L \right\}, \quad (1)$$

where $L(i)$ denotes the ith data in L, and $L\_\text{count}(i)$ denotes number of nodes from which the average value is calculated.

Here we demonstrate an example to illustrate the two basic operations. Let $L = \{0, 36.1, 34.5, 35.3\}$, and $L\_\text{count} = \{0, 1, 1, 2\}$. Then, $L + 36.9 = \{(0 \times 0 + 36.9)/(0 + 1), (36.1 \times 1 + 36.9)/(1 + 1), (34.5 \times 1 + 36.9)/(1 + 1), (35.3 \times 2 + 36.9)/(2 + 1)\} = \{36.9, 36.5, 35.7, 35.83\}$. And merge $L[L, L + 36.9] = \{0, 36.1, 34.5, 35.3\} + \{36.9, 36.5, 35.7, 35.83\} = \{0, 34.5, 36.1, 35.3, 36.9, 36.5, 35.7, 35.83\}$, $L\_\text{count} = \{0, 1, 1, 2\} + \{1, 2, 2, 3\} = \{0, 1, 1, 2, 1, 2, 2, 3\}$.

In the following we illustrate the CNS computing process for s_3 in Figure 1(c) by assuming that the residual energy is identical to all. The input for the CNS computing problem is described as $CS_3 = \{s_1, s_3, s_7, s_8\}$, ANS $- \{s_3\} = \{s_5, s_{10}, s_{13}\}$, and $X = \{36.1, 34.5, 36.9\}$. Initially, $L = \{0\}$, $L\_\text{count} = \{0\}$, and the corresponding set list as $\{\{\emptyset\}\}$.

(1) Consider the sensing data 36.5 of s_5 : $L = \{0, 36.1\}$, $L\_\text{count} = \{0, 1\}$, and the corresponding set list as $\{\{\emptyset\}, \{s_5\}\}$;

(2) consider the sensing data 34.5 of s_{10} : $L = \{0, 36.1\} + \{34.5, 35.3\} = \{0, 36.1, 34.5, 35.3\}$, $L\_\text{count} = \{0, 1, 1, 2\}$, and the corresponding set list as $\{\{\emptyset\}, \{s_5\}, \{s_{10}\}, \{s_5, s_{10}\}\}$;

(3) consider the sensing data 36.9 of s_{13} : $L = \{0, 36.1, 34.5, 35.3\} + \{36.9, 36.5, 35.7, 35.83\} = \{0, 34.5, 36.1, 35.3, 36.9, 36.5, 35.7, 35.83\}$, $L\_\text{count} = \{0, 1, 1, 2, 1, 2, 2, 3\}$, and the set list as $\{\{\emptyset\}, \{s_5\}, \{s_{10}\}, \{s_5, s_{10}\}, \{s_{13}\}, \{s_5, s_{13}\}, \{s_{10}, s_{13}\}, \{s_5, s_{10}, s_{13}\}\}$.

Input: $G = (V, E)$, ε, $X = \{x_1, x_2, \ldots, x_n\}$, Energy $= \{e_1, e_2, \ldots, e_n\}$;
Output: ANS, CS.

(1) //Initialization_Process ( )
(2) Calculate DCR $= \{\mathrm{DCR}_1, \mathrm{DCR}_2, \ldots, \mathrm{DCR}_n\}$
(3) Set the state of all nodes as *Un-Covered*;
(4) Sort nodes into sequence with partially ordered tuple ⟨data coverage range, residual energy⟩;
(5) ANS $\leftarrow \emptyset$, CS $\leftarrow \emptyset$;
(6) **for** one maximal i in the sequnce with state as *Un-Covered*
(7) ANS $\leftarrow \{i\}$ + ANS, and set i as *Primary-Covered*;
(8) **for** each $j \in \mathrm{DCR}_i - \mathrm{ANS}$ with state as *Un-Covered*
(9) $\mathrm{CS}_i \leftarrow \{j\} + \mathrm{CS}_i$, and set j as *Primary-Covered*;
(10) **end for**
(11) **end for**
(12) //Cover_Set_Balance_Process ( )
(13) Sort CS into a sequence with decreasing order of the set size;
(14) **for** each CS_i in the sequence
(15) Sort nodes in CS_i with decreasing order of their deviation to i;
(16) **for** each j in the sequence
(17) find out all k which satisfies $j \in \mathrm{CS}_k$ and $|\mathrm{CS}_k| < |\mathrm{CS}_i|$, select k with minimal cover set size;
(18) $\mathrm{CS}_i \leftarrow \mathrm{CS}_i - \{j\}$, $\mathrm{CS}_k \leftarrow \mathrm{CS}_k + \{j\}$;
(19) **end for**
(20) **end for**
(21) //Node_Replace_Process ( )
(22) **for** each $i \in \mathrm{ANS}$
(23) **for** each $j \in \mathrm{CS}_i$
(24) if $d(x_j, x_k) \leq \varepsilon$ for any $k \in \mathrm{CS}_i + \{i\} - \{j\}$, then mark j as a candidate of i;
(25) **end for**
(26) select a node m from all candidates of i with maximal residual energy e_i;
(27) ANS $\leftarrow$ ANS + $\{m\} - \{i\}$; $\mathrm{CS}_m = \mathrm{CS}_i + \{i\} - \{m\}$;
(28) **end for**

Algorithm 1: Pseudocodes for Cover Set Balance (CSB) algorithm.

The deviation between 35.3 and the sensing data of nodes in set $\mathrm{CS}_3 + \{s_3\}$ is no more than 0.5, and it is similar to 35.7. Accordingly, the corresponding correlated node sets are $\{s_5, s_{10}\}$ and $\{s_5, s_{10}, s_{13}\}$ located at the 4th and 7th positions in L. Finally, $\mathrm{CNS}_3 = \{s_5, s_{10}\}$ followed by $|\{s_5, s_{10}\}| < |\{s_5, s_{10}, s_{13}\}|$.

Algorithm 2 provides the pseudocodes for CNSC algorithm.

3.2.2. HREF Algorithm. For a given $i \in \mathrm{ANS}$, its sensing data is computed with the nodes in CNS_i, which makes it possible to shut off to preserve energy. The basic idea of the HREF algorithm is described as follows: (1) build the active node set ANS with CSB algorithm; (2) for each $i \in \mathrm{ANS}$, calculate its correlated node set CNS_i; (3) remove certain active nodes from ANS. The pseudocodes are shown in Algorithm 3.

In Line 3, an active node set is generated with respect to the concept of data coverage range and corresponding correlated node set in ANS. Then we mark all active nodes as *Un-Completed*. There are two different states for each node in the active node set, namely, *Completed* and *Un-Completed*. In Line 4, we sort CNS with ascending order of their set size. In Line 5–10, we check whether if an active node can be removed from ANS and mark each node in CNS_j as *Completed*.

4. Theoretical Analysis

Theorem 7. *The CSB and HREF algorithms correctly generate an active node set for a given wireless sensor network even in case that there are message losses.*

Proof. The cases with CSB and HREF are described as follows.

(1) Firstly, we prove that the sink node obtains all sensing data of the nodes in *Closed state* through the selected active node set. At the beginning of CSB and HREF, all nodes are active nodes. The state that whether one node is closed or not depending on the condition whether the sensing data can be fused by the corresponding correlated node set. In these algorithms, the node is removed from the active node set only in case the condition is satisfied. Thus it is sure that all sensing data can be obtained from nodes in ANS calculated via CSB and HREF.

(2) Secondly, we prove that CSB and HREF correctly generate an active node set even in case that there are message losses. Note that our algorithms aim at shutting down certain nodes if they can be fused by other active nodes, which means that these nodes keep active if the above condition is not satisfied. It is obvious that the message losses never reduce the

Input: ANS, ε, CS = $\{CS_1, CS_2, \ldots, CS_n\}$, $X = \{x_1, x_2, \ldots, x_n\}$, Energy = $\{e_1, e_2, \ldots, e_n\}$;
Output: ANS.

(1) **for** each i in ANS
(2) $CNS_i = \emptyset$;
(3) **for** each *maxd* nodes in ANS
(4) placed them in *node_vector*;
(5) $L[0] = \{0\}$;
(6) **for** $j = 1$ to $|node\_vector|$, $L = mergeL\ (L, L + x_{node\_vector(j)})$;
(7) **for** each l in $L[i]$
(8) **for each**$k \in CS_i + \{i\}$ if $|l - x_k| \leq \varepsilon$ **then** $temp = L\_pos(l)$;
(9) **for**$k = |node\_vector| - 1$ to 0
(10) **if** $temp > 2^k$ and $temp \leq 2^{(k+1)}$
(11) $Dset = Dset + \{$the kth node in *node_vector*$\}$, $temp = temp - 2^k$;
(12) **end if**
(13) **end for**
(14) if $(|Dset| < |CNS_i|)$ or $(|Dset| = |CNS_i|$ and $\hat{e}(|Dset|) > \hat{e}(|Dset|)$, then $CNS_i \leftarrow Dset$;
(15) **end for**
(16) **end for**
(17) **end for**

Algorithm 2: Pseudocodes for CNSC algorithm.

Input: $G = (V, E)$, ε, $X = \{x_1, x_2, \ldots, x_n\}$, Energy = $\{e_1, e_2, \ldots, e_n\}$;
Output: ANS.

(1) Run CSB algorithm to obtain the initial ANS and CS;
(2) Run CNSC algorithm to obtain CNS;
(3) Mark all nodes in ANS as *Un-Completed*;
(4) Sort CNS with increasing order of their set size;
(5) **for** one minimal CNS_i in the sequence with $CNS_i \subseteq ANS$
(6) **if** $CNS_i \neq \emptyset$, and the state of i is *Un-Completed*, then
(7) ANS $\leftarrow$ ANS $- \{i\}$;
(8) **for** each $j \in CNS_i$, mark j as *Completed*;
(9) **end if**
(10) **end for**

Algorithm 3: Pseudocodes for HREF algorithm.

number of active nodes, and thus CSB and HREF correctly generate an active node set correctly in case of message losses. □

Theorem 8. *The active node set size with CSB is at most* $(1 + \log n) \times |OPT1|$*, where* $OPT1$ *is the optimal active node set with respect to the concept of data coverage range and* n *is the number of nodes in sensor network.*

Proof. The active node selection problem with respect to the concept of data coverage range is essentially a set covering problems [27]. We regard the problem of selecting a smallest size of active node set as the problem of selecting the minimum size of subset in set-covering issue [22]. Similar to the greedy approximation algorithm of set covering problem, CSB also takes the greedy strategy to maximize the size of data coverage range for each new added active node. Let δ be the size of selected active node set with number of nodes |OPT1|, and let $\{DCR_{\delta 1}, DCR_{\delta 2}, \ldots, DCR_{\delta |OPT1|}\}$ be the corresponding data coverage ranges of each active node. For each data coverage range $DCR_{\delta i}$, the maximal number of selected active nodes is at most $(1 + \log(|DCR_{\delta i}|))$ with the above greedy strategy. The total number of selected active nodes is

$$N_1 \leq \sum_{1}^{|OPT1|} \left(1 + \log\left(\left|CR_{\delta_i}\right|\right)\right) \leq |OPT1| \times \left(1 + \log\left(\max\_CR\right)\right), \tag{2}$$

where $\max\_DCR = \max\{|DCR_{\delta i}| \mid i \in V\}$.

Due to $\max\_DCR \leq n$, the size of the active node set with CSB is at most $(1 + \log n) \times |OPT1|$. □

Theorem 9. *The time complexity of CSB is* $O(n^2)$.

Proof. The CSB algorithm is divided into three processes as mentioned.

In the Initialization_Process, it is easy to know that the time complexity of obtaining all node's data coverage range is $O(n^2)$. The time complexity of sorting nodes with the partially ordered tuple is $O(n \log n)$. The time complexity of selecting a node with maximal data coverage range in the sequence is $O(n)$ and the process runs $O(n)$ times. So the time complexity of Initialization_Process is $O(n^2)$.

In the Cover_Set_Balance_Process, the time complexity for each covered node to find the active node is $(n - |\text{ANS}|) \times (|\text{ANS}| - 1)$, where $(n - |\text{ANS}|)$ denotes the number of covered nodes and $|\text{ANS}|$ denotes the number of active nodes. So the time complexity of the process is $O(n^2)$.

In the Node_Replace_Process, the progress of selecting the optimized candidate active node and replacing the low-energy node is carried out simultaneously, and the time complexity is $O(n)$.

So the time complexity of CSB is $O(n^2)$. □

Theorem 10. *The size of the active node set with HREF is at most* $(1 + \log((1 + \log n) \times |OPT1|)) \times |OPT2|$, *where* $OPT2$ *is the optimal active node set and n is the number of nodes in the network.*

Proof. We adopt a greedy strategy HREF to solve the active node selection problem. The HREF is divided into two phrases: the first step is the CSB algorithm and the second phrase is to further reduce the number of active nodes selected by CSB.

Assume that the size of active node set with CSB is m. According to [28], the optimized number of active nodes has upper bound as $(1 + \log m) \times |\text{OPT2}|$, where OPT2 is the optimal node set with respect to the concept of correlated node set and depends on OPT1. According to Theorem 8, the maximal number of active nodes selected by CSB is $m \leq (1 + \log n) \times |\text{OPT1}|$. Then the upper bound for the number of active nodes selected by HREF is $(1 + \log((1 + \log n) \times |\text{OPT1}|)) \times |\text{OPT2}|$. □

Theorem 11. *The time complexity of HREF is* $O(n^2 + m \times \binom{m}{maxd} \times 2^{maxd} + m^2)$, *where* $m = (1 + \log n) \times |OPT1|$ *is the maximal number of active nodes selected by CSB.*

Proof. The time complexity of HREF includes three different phases: the first step runs the CSB algorithm, the second step runs the CNSC algorithm, and the third step shuts down certain nodes. The time complexity for the first step is discussed above as $O(n^2)$. In the second step, each node $i \in \text{ANS}$ spends time $O(\binom{m}{maxd} \times 2^{maxd})$ to compute an optimized correlated node set from all its correlated node sets, where $\binom{m}{maxd}$ denotes the number of subsets and 2^{maxd} denotes the time complexity of the sequence L. So all nodes totally cost $O(m \times \binom{m}{maxd} \times 2^{maxd})$ to calculate their corresponding correlated node set. In the third step, the process of shutting down redundant active nodes runs $O(m^2)$ times. Thus, the total time complexity of HREF is $O(n^2 + m \times \binom{m}{maxd} \times 2^{maxd} + m^2)$. □

5. Simulation Results and Analysis

In this section, we demonstrate detailed simulation experiments to evaluate the actual performance of the above algorithms. Note that this paper focuses on the active node selection problem by exploiting correlations among nodes but has no concern with the aggregation operators or probabilistic models. We compare the proposed CSB and HREF algorithms with the DClocal, DCglobal [22], EEDC [24], and Snapshot [28] by running them in the same networks as well as the same parameters for the environment.

Here we adopt two main metrics for the algorithm performance, namely, the number of active nodes and the network lifetime. The number of active nodes is an important measurement since data coverage basically aims at minimizing the number of active nodes. We compare the related algorithms via this metric for a given data collection epoch. Meanwhile, the active node selection problem aims at maximizing the network lifetime, and thus network lifetime is adopted as the other metric for the performance comparison.

In this section, we first introduce the simulation environment, then compare the algorithms via the number of active nodes with different parameters, such as network size, error threshold, and number of events, and finally we compare them by the metric of network lifetime with different parameters as well as interval for each epoch.

5.1. Simulation Environment Setup. We adopt MATLAB as the platform tool which is popularly used in the simulation of wireless sensor networks. The network is set up by placing $|V|$ nodes in a random manner. The events are randomly deployed in the monitored region. The cost of information collection is assumed 0.1 units during each epoch.

We adopt the approach of generating synthetic sensor data on the monitored region. In the synthetic data set, h events are randomly generated as $\text{Event} = \{\text{Event}_1(t), \text{Event}_2(t), \ldots, \text{Event}_h(t)\}$ and they are also randomly deployed in the monitored region. The sensing data for a given node is affected by these events which is inversely proportional to their distance. The initial data of each event is randomly selected from [20, 40]. The value of an event Event_i at time t is formulated as $\text{Event}_i(t) = \text{Event}_i(t - interval) + Z$ where Z is a random variable that follows the normal distribution with mean 0 and variance 0.1, while $\text{Event}_i(0)$ is the initial value of the ith event. The data of node s at time t is computed by Formula (3):

$$x_s(t) = \sum_{i=1}^{m} \frac{1/\left(\text{dist}\left(s, \text{Event}_i\right)\right)}{\sum_{j=1}^{m} \left(1/\left(\text{dist}\left(s, \text{Event}_j\right)\right)\right)} \times \text{Event}_i(t), \quad (3)$$

where $\text{dist}(s, \text{Event}_i)$ denotes the square of the distance between node s and event Event_i and h denotes the number of events.

Table 1: Default values for the simulation parameters.

Parameter description	Default value
Target area size	100 m × 100 m
Network size	200
The location of sink	(50, 50)
Transmission radius	20 m
Number of events	10
Error threshold	0.5
maxd	8
Initial energy of each node	100 units
Energy cost for sensing during each epoch	0.02 units
Energy cost for transmission during each epoch	0.03 units
Fraction of alive nodes	75%
Interval for reselecting a new active node set	80 epochs

In this paper we focus on the node selection process and its impact on the network lifetime, while the routing/path selection are both ignored. Readers are guided to other works for details about these issues [29–31]. The default values for the simulation parameters are listed in Table 1.

5.2. Comparison of Number of Active Nodes. In this part, we compare the performance of our algorithms with related works by various parameters, including network size, error threshold, and the number of events.

5.2.1. Impact of Network Size. The network size is set from 100 to 500 with increment as 100, and the simulation result is demonstrated in Figure 2. It shows that the number of the selected active nodes ascends with the network size when the network size is smaller than 400. However, this trend is not obvious when the network size is large enough ($n = 500$). A certain number of active nodes are selected to perform the data collection process especially when the network is dense enough. This trend demonstrates the importance of active node selection with correlative optimization during the data collection process.

HREF always has better performance compared with CSB, as we can see from Figure 2. For example, the number of active nodes selected by HREF is only 80.91% of that by CSB in case that the network size is 300. It demonstrates that HREF is rather significant to reduce the active nodes by removing nodes which can be computed by the corresponding correlated node set with the help of CNSC algorithm.

In all cases, HREF and CSB have better performance compared with related algorithms, that is, EEDC, DCglobal, Snapshot, and DClocal. When $n = 300$, the number of selected node is 15.05, 18.6, 20.75, 30.15, 28.35, and 34.65 with HREF, CSB, DCglobal, EEDC, Snapshot, and DClocal.

5.2.2. Impact of Error Threshold. The error threshold varies from 0.1 to 1.15 with increment as 0.15 in the simulations. As shown in Figure 3, the number of active nodes selected by

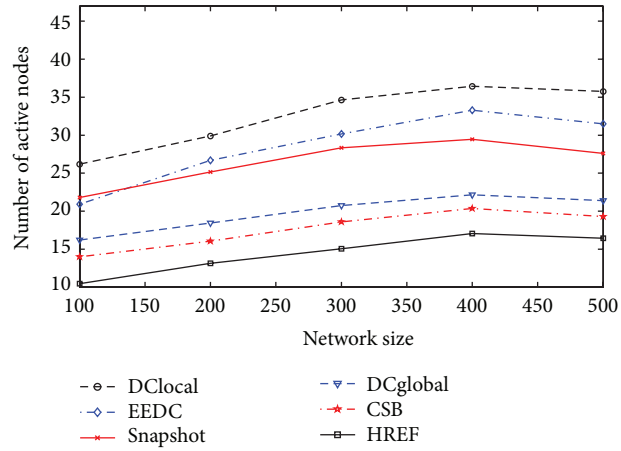


Figure 2: The impact of network size on the number of active nodes.

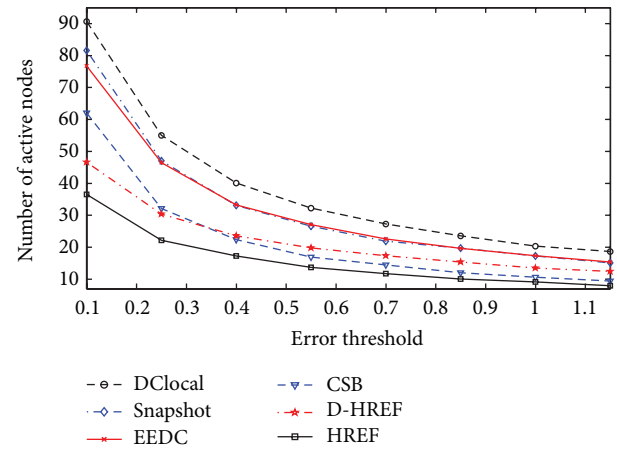


Figure 3: The impact of error threshold on the number of active nodes.

HREF is lower than other algorithms in all cases. As the error threshold increases in the range of [0.1, 0.55], the number of active nodes decreases significantly. However, it is not obvious in the case that the error threshold is larger than 0.7. Hence, it is helpful to reduce the number of active nodes if a larger error threshold is tolerant in some applications.

5.2.3. Impact of the Number of Events. The number of events varies from 5 to 40 with increment as 5 and the simulation result is demonstrated in Figure 4. It shows that the number of selected active nodes is independent of the number of events by using the data computing Formula (2). It can be seen that HREF and CSB have better performance compared with related algorithms regardless of the number of events.

5.3. Comparison of Network Lifetime. There are variations of measurement for network lifetime [27], such as the first node to die, the number of alive nodes, and the fraction of alive nodes. The measurement with the first node to die is not a good measure metric in practical applications, especially in the dense-deployed wireless sensor networks.

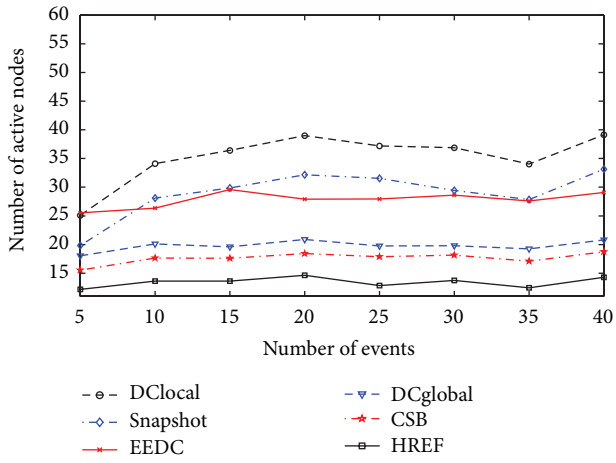


Figure 4: The impact of the number of events on the number of active nodes.

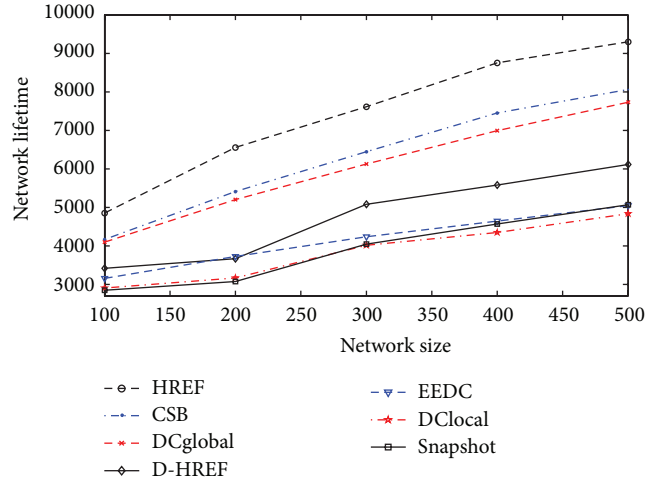


Figure 5: The impact of network size on the network lifetime.

This is because the redundancy among correlated nodes is helpful to illuminate the defect of single-node failure. The definition based on fraction of alive nodes regards that the network is alive when the fraction of surviving nodes remains above a given threshold [32]. The network lifetime is defined in this paper as the time period during which the fraction of alive nodes remains above a given threshold and they are also connected.

To measure the network lifetime, we have to determine the relay nodes forwarding the sensing data from active nodes by constructing a minimum Steiner tree [33]. The nodes selected by the minimum Steiner tree construction step are called Steiner nodes. Note that the relay nodes do not need to sense data. In the following experiments, we compared the network lifetime of our algorithms to related algorithms in various environmental parameters.

5.3.1. Impact of Network Size. The network size is set from 100 to 500 with increment as 100, and the simulation result is demonstrated in Figure 5. It shows that the network lifetime increases along with the network size increasing. This is reasonable because the number of selected nodes might be independent on the network size. When there is enough data redundancy among the sensing data, more redundant nodes are used to extend the network lifetime, as shown in Section 5.2.1, HREF and CSB have better performance compared with related algorithms regardless of network size. Especially, our algorithm works better when the network size is larger than 200.

The HREF has significant improvement on the network lifetime compared with CSB too. For example, the lifetime has about 18.19% increment compared with CSB in case that the network size is 300. It is reasonable since we adopt not only node reduction but also node replacement strategies which are rather helpful to enlarge the network lifetime.

5.3.2. Impact of Error Threshold. The error threshold varies from 0.1 to 1.15 with increment as 0.15 in the simulations.

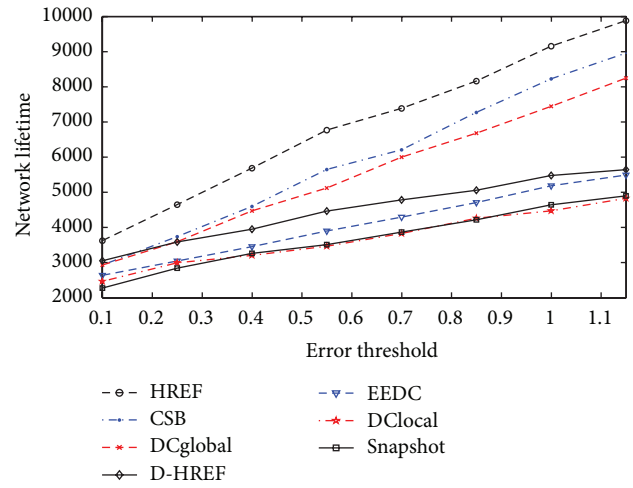


Figure 6: The impact of error threshold on the network lifetime.

As shown in Figure 6, the network lifetime increases along with the error threshold increasing. HREF and CSB have better performance compared with the related algorithms, that is, EEDC and DCglobal. CSB has a better performance compared with DCglobal especially when the error threshold is larger than 0.4. The network lifetime of HREF algorithm is longer than the other algorithms in all cases.

5.3.3. Impact of Interval. The value of *interval* varies from 20 to 160 with increment as 20 in the simulations. In Figure 7, the network lifetime increases along with the *interval* when it is smaller than 80. However, this trend slows down when *interval* is large than 80. It means that it benefits to extend the network lifetime if a larger *interval* is tolerant in some applications. In addition, HREF and CSB have better performance compared with related algorithms regardless of the *interval*.

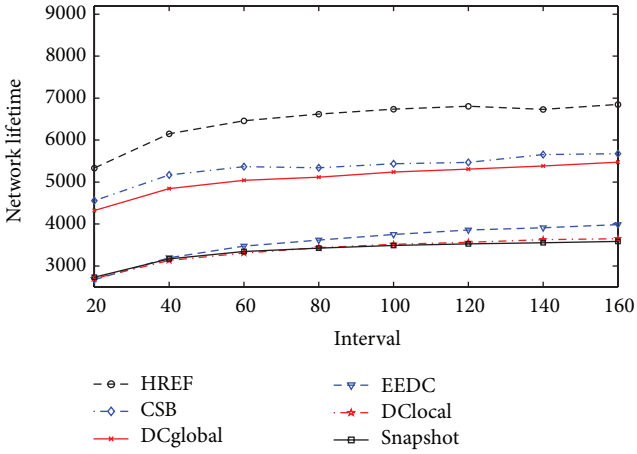


Figure 7: The impact of *interval* on the network lifetime.

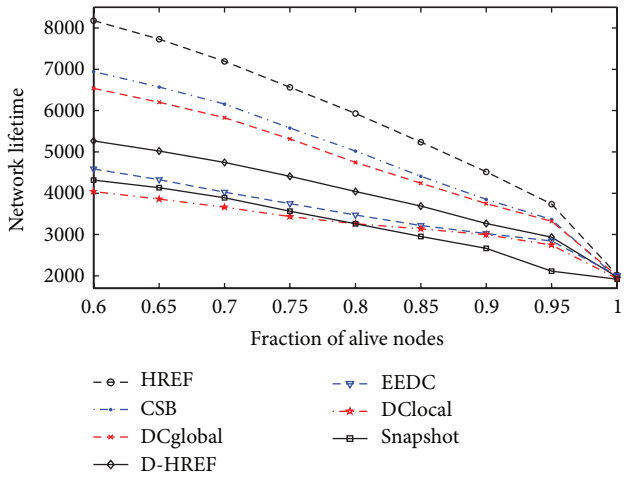


Figure 8: The impact of the fraction of alive nodes on the network lifetime.

5.3.4. Impact of Fraction of Alive Nodes. The fraction of alive nodes varies from 0.6 to 1 with increment as 0.05 in the simulations. In Figure 8, the network lifetime decreases along with the fraction of alive nodes increasing. The HREF has better performance compared with related algorithms. The network lifetime of CSB is longer than that of DCglobal when the fraction of alive nodes is smaller than 0.95. However the case changes when the fraction of alive nodes is larger than 0.95. This is because CSB balances between the data coverage range priority and the energy priority. As the data coverage range priority prefers to select nodes with larger data coverage ranges, these nodes with lower energy might be selected as well, which results in rapid node failure and a dying network. The similar conclusion is drawn in Section 3.1. However, as the measurement of the first node to die is not suitable measure metric for network lifetime evaluation in practical applications, the CSB is still better than DCglobal in this case.

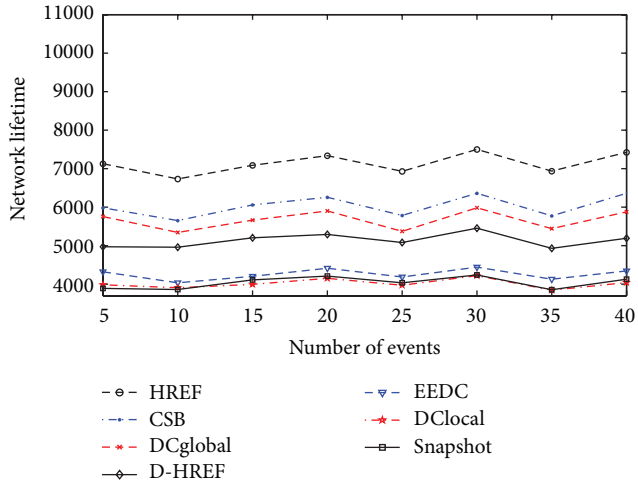


Figure 9: The impact of the number of events on the network lifetime.

5.3.5. Impact of Number of Events. The number of events varies from 5 to 40 with increment as 5 and the simulation result is demonstrated in Figure 9. It shows that network lifetime is independent of the number of events. However, HREF and CSB have better performance compared with related algorithms regardless of the number of events.

6. Related Works

Energy efficiency is a critical design consideration in battery powered and densely deployed wireless sensor networks, which can be achieved by minimizing the number of messages transmitted during the data collection process. Related works include clustering, network coding, in-network data aggregation, and approximate data collection.

Clustering is proven to be an effective approach to provide better data aggregation and scalability for large wireless sensor network [6–11]. Recently, Aslam et al. [7] propose a novel multicriterion optimization technique based on energy-efficient clustering approach. This method takes multiple individual metrics as inputs in the cluster head selection process and simultaneously optimizes the energy efficiency of each individual node as well as the overall system. Karaboga et al. [8] propose an energy-efficient clustering mechanism based on artificial bee colony algorithm to prolong the network lifetime. The simulation results show that the artificial bee colony algorithm based clustering approach can be applied to routing protocols successfully. Naeimi et al. [9] classify routing protocols according to their different objectives and methods by addressing both the shortcomings and the strength of clustering process on each stage of cluster head selection, cluster formation, data aggregation, and data communication and summarized them into categories. Moreover, Lloret et al. demonstrated in [10] that cluster-based mechanisms allow multiple types of network topologies in order to have the most efficient network. Lehasini et al. [11] used clusters to improve the network coverage.

In-network data aggregation [12–18] is another approach to reduce the amount of data transmitted by the nodes and prolong the network lifetime. It performs data aggregation in network to reduce the amount of data transmission by constructing a routing tree. In [12, 13] we can find complete surveys on distributed database management techniques and data aggregation for wireless sensor networks. Al-Karaki et al. [14] present a Grid-based Routing and Aggregator Selection Scheme (GRASS), which achieves low-energy dissipation and low-latency without sacrificing quality. Seyin et al. [15] propose a localized and energy-efficient data aggregation tree approach called Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs) for sensor networks. Gao et al. [16] jointly adopt the cooperative multiple-input-multiple-output and data-aggregation techniques to reduce the energy consumption per bit in wireless sensor network by reducing the amount of data for transmission and better using network resources through cooperative communication.

Approximate data collection is also an energy-efficient approach which is further divided into two subcategories. The first subcategory is approximate data collection via probabilistic models of sensing data collected from wireless sensor networks [19, 20]. Xua and Choi [19] propose a new class of Gaussian processes for resource-constrained mobile sensor networks and propose a distributed algorithm which achieves the field prediction by correctly fusing all observations. Min and Chung [20] present an approximate data gathering approach which utilizes temporal and spatial correlations for wireless sensor network and does not transmit the data to the sink if the data are accurately predicted. The second subcategory is approximate data gathering without probabilistic models. Kotidis [23] propose Snapshot queries for energy-efficient data acquisition in sensor networks. They constitute a network Snapshot through selecting a set of active nodes which is used to provide quick approximate answers to user queries and reducing the energy consumption substantially in wireless sensor network. Gupta et al. [28] design techniques that exploit data correlation among nodes to minimize communication costs incurred during data gathering in a wireless sensor network. They design distributed algorithms that can be implemented in an asynchronous communication model. They also design an exponential approximation algorithm that returns a solution within $O(\log n)$ of the optimal size. Liu et al. [24] propose a data collection approach based on a careful analysis of the sensor data. By exploring the spatial correlation of sensing data, they dynamically divide the nodes into clusters such that the sensors in the same cluster have similar sensing time series which can share the workload of data collection since their future data may likely be similar. Hung et al. [22] propose an algorithm to determine a set of active nodes with high residual energy and wide data coverage ranges. Here, the data coverage range of a node is the set of nodes that have sensor data very close to the particular node. They also develop an algorithm to further reduce the extra cost incurred in messages collection and transmission for selection of active nodes.

In previous work, we have studied the minimum-latency data aggregation problem and proposed a new efficient scheme for it [34]. The basic idea is that we first build an aggregation tree by ordering nodes into layers and then we proposed a scheduling algorithm on the basis of the aggregation tree to determine the transmission time slots for all nodes in the network with collision avoiding. We have proved that the upper bound for data aggregation with our proposed scheme is bounded by $(15R + \Delta - 15)$ for wireless sensor networks in two-dimensional space, where Δ is the maximum degree and R is the network radius. We have also simulated the case in three-dimensional wireless sensor networks and proposed an aggregation tree construction algorithm based on maximum independent set [35]; the height of the spanning tree can be reduced to about 50%.

In previous work, we study the node selection problem with data accuracy guaranteed in service-oriented wireless sensor networks [36]. We exploit the spatial correlation between the service data and aim at selecting minimum number of nodes to provide services with data accuracy guaranteed. Firstly, we have formulated this problem into an integer nonlinear programming problem to illustrate its NP-hard property. Secondarily, we have proposed two heuristic algorithms, namely, Separate Selection Algorithm (SSA) and Combined Selection Algorithm (CSA). The SSA is designed to select nodes for each service in a separate way, and the CSA is designed to select nodes according to their contribution increment.

7. Conclusions

Due to the correlation and redundancy among the sensing data in wireless sensor networks, it is an important issue to develop an energy-efficient active node selection strategy, which not only improves the network lifetime but also is helpful to solve other problems, such as lower network throughput and serious node conflict in dense wireless sensor networks. In this paper, we concern with the active node selection issue and provided a formal definition for this problem. We propose the Cover Sets Balance (CSB) algorithm and High Residual Energy First nodes selection (HREF) algorithm aiming at extending the network lifetime of wireless sensor networks. We also propose a Correlated Node Set Computing (CNSC) algorithm to find the correlated node set for a given node. Experimental results on synthesized data sets show that HREF can significantly reduce the number of active nodes, and these algorithms are able to significantly extend the network lifetime compared with related works. In the future work, we are to further consider the temporal correlation among the sensing data and design an efficient node scheduling scheme with both spatial and temporal correlation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Science Foundation of China under Grand nos. 61370210 and 61103175, Fujian

Provincial Natural Science Foundation of China under Grant nos. 2011J01345, 2013J01232, and 2013J01229, and the Development Foundation of Educational Committee of Fujian Province under Grand no. 2012JA12027. It has also been partially supported by the "Ministerio de Ciencia e Innovación," through the "Plan Nacional de I+D+i 2008–2011" in the "Subprograma de Proyectos de Investigación Fundamental," Project TEC2011-27516, and by the Polytechnic University of Valencia, though the PAID-15-11 multidisciplinary Projects.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] S. Sendra, J. Lloret, M. García, and J. F. Toledo, "Power saving and energy optimization techniques for wireless sensor networks," *Journal of Communications*, vol. 6, no. 6, pp. 439–459, 2011.

[3] O. Diallo, J. Rodrigues, M. Sene, and J. Lloret, "Distributed database management techniques for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, 2013.

[4] L. M. L. Oliveira, J. J. P. C. Rodrigues, A. G. F. Elias, and B. B. Zarpelão, "Ubiquitous monitoring solution for Wireless Sensor Networks with push notifications and end-to-end connectivity," *Mobile Information Systems*, vol. 10, no. 1, pp. 19–35, 2014.

[5] O. Diallo, J. J. P. C. Rodrigues, and M. Sene, "Real-time data management on wireless sensor networks: a survey," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 1013–1021, 2012.

[6] O. Boyinbode, H. Le, and M. Takizawa, "A survey on clustering algorithms for wireless sensor networks," *International Journal of Space-Based and Situated Computing*, vol. 1, no. 2, pp. 130–136, 2011.

[7] N. Aslam, W. Phillips, W. Robertson, and S. Sivakumar, "A multi-criterion optimization technique for energy efficient cluster formation in wireless sensor networks," *Information Fusion*, vol. 12, no. 3, pp. 202–212, 2011.

[8] D. Karaboga, S. Okdem, and C. Ozturk, "Cluster based wireless sensor network routing using artificial bee colony algorithm," *Wireless Networks*, vol. 18, no. 7, pp. 847–860, 2012.

[9] S. Naeimi, H. Ghafghazi, C. O. Chow, and H. Ishii, "A survey on the taxonomy of cluster-based routing protocols for homogeneous wireless sensor networks," *Sensor*, vol. 12, no. 6, pp. 7350–7409, 2012.

[10] J. Lloret, M. Garcia, D. Bri, and J. R. Diaz, "A cluster-based architecture to structure the topology of parallel wireless sensor networks," *Sensors*, vol. 9, no. 12, pp. 10513–10544, 2009.

[11] M. Lehasini, H. Guyennet, and M. Feham, "Cluster-based energy-efficient k-coverage for wireless sensor networks," *Network Protocols and Algorithms*, vol. 2, no. 2, pp. 89–106, 2010.

[12] R. Rajagopalan and P. K. Varshney, "Data aggregation techniques in sensor networks: a survey," *IEEE Communications Surveys*, vol. 6, no. 4, pp. 48–63, 2006.

[13] K. Maraiya, K. Kant, and N. Gupta, "Wireless sensor network: a review on data aggregation," *International Journal of Scientific & Engineering Research*, vol. 2, no. 4, pp. 1–6, 2011.

[14] J. N. Al-Karaki, R. Ul-Mustafa, and A. E. Kamal, "Data aggregation and routing in Wireless Sensor Networks: optimal and heuristic algorithms," *Computer Networks*, vol. 53, no. 7, pp. 945–960, 2009.

[15] H. O. Tan, I. Korpeoglu, and I. Stojmenovic, "Computing localized power-efficient data aggregation trees for sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 3, pp. 489–500, 2011.

[16] Q. Gao, Y. Zuo, J. Zhang, and X.-H. Peng, "Improving energy efficiency in a wireless sensor network by combining cooperative MIMO with data aggregation," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 8, pp. 3956–3965, 2010.

[17] G. Wei, Y. Ling, B. Guo, B. Xiao, and A. V. Vasilakos, "Prediction-based data aggregation in wireless sensor networks: combining grey model and Kalman Filter," *Computer Communications*, vol. 34, no. 6, pp. 793–802, 2011.

[18] L. Xiang, J. Luo, and A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," in *Proceedings of the 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '11)*, pp. 46–54, Salt Lake City, Utah, USA, June 2011.

[19] Y. Xua and J. Choi, "Spatial prediction with mobile sensor networks using gaussian processes with built-in gaussian markov random fields," *Automatica*, vol. 48, no. 8, pp. 1735–1740, 2012.

[20] J.-K. Min and C.-W. Chung, "EDGES: efficient data gathering in sensor networks using temporal and spatial correlations," *Journal of Systems and Software*, vol. 83, no. 2, pp. 271–282, 2010.

[21] J. Li and S. Cheng, "(ε, δ)-Approximate aggregation algorithms in dynamic sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 3, pp. 385–396, 2012.

[22] C. C. Hung, W. C. Peng, and W. C. Lee, "Energy-aware set-covering approaches for approximate data collection in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 1993–1200, 2012.

[23] Y. Kotidis, "Snapshot queries: towards data-centric sensor networks," in *Proceedings of the 21st International Conference on Data Engineering (ICDE '05)*, pp. 131–142, April 2005.

[24] C. Liu, K. Wu, and J. Pei, "An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 7, pp. 1010–1023, 2007.

[25] X. Zhang, H. Wang, F. Naït-Abdesselam, and A. A. Khokhar, "Distortion analysis for real-time data collection of spatially temporally correlated data fields in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1583–1594, 2009.

[26] E. Karasabun, I. Korpeoglu, and C. Aykanat, "Active node determination for correlated data gathering in wireless sensor networks," *Computer Networks*, vol. 57, no. 5, pp. 1124–1138, 2013.

[27] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction To Algorithms*, McGraw Hill, 2001.

[28] H. Gupta, V. Navda, S. Das, and V. Chowdhary, "Efficient gathering of correlated data in sensor networks," *ACM Transactions on Sensor Networks*, vol. 4, no. 1, article 4, pp. 402–413, 2008.

[29] G. Campobello, A. Leonardi, and S. Palazzo, "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, pp. 191–205, 2012.

[30] L. C. Tseng, F. T. Chien, D. Zhang, R. Y. Chang, W. H. Chung, and C. Y. Huang, "Network selection in cognitive heterogeneous networks using stochastic learning," *IEEE Communications Letters*, vol. 17, no. 12, pp. 2304–2307, 2013.

[31] J. J. P. C. Rodrigues and P. A. C. S. Neves, "A survey on IP-based wireless sensor network solutions," *International Journal of Communication Systems*, vol. 23, no. 8, pp. 963–981, 2010.

[32] A. A. Aziz, Y. A. Sekercioglu, P. Fitzpatrick, and M. Ivanovich, "A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 121–144, 2012.

[33] K. Mehlhorn, "A faster approximation algorithm for the Steiner problem in graphs," *Information Processing Letters*, vol. 27, no. 3, pp. 125–128, 1988.

[34] C. Hongju, L. Qin, and J. Xiaohua, "Heuristic algorithms for real-time data aggregation in wireless sensor networks," in *Proceedings of the International Conference on Wireless Communications and Mobile Computing (IWCMC '06)*, pp. 1123–1128, Vancouver, Canada, July 2006.

[35] F. Li and H. Cheng, "An efficient scheme for minimum-latency data aggregation in two- and three-dimensional wireless sensor networks," in *Proceeding of the 2nd International Conference on Cloud and Green Computing (CGC '12)*, pp. 252–259, Xiangtan, China, 2012.

[36] H. Cheng, R. Guo, and Y. Chen, "Node selection algorithms with data accuracy guarantee in service-oriented wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 527965, 14 pages, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 393248, 12 pages
http://dx.doi.org/10.1155/2014/393248

Review Article

Securing Cognitive Wireless Sensor Networks: A Survey

Alexandros Fragkiadakis,[1] Vangelis Angelakis,[2] and Elias Z. Tragos[1]

[1] *Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion, 71110 Crete, Greece*
[2] *Department of Science and Technology, Linköping University, 58183 Linköping, Sweden*

Correspondence should be addressed to Alexandros Fragkiadakis; alfrag@ics.forth.gr

Received 1 January 2014; Accepted 23 February 2014; Published 27 March 2014

Academic Editor: Christos Verikoukis

Wireless sensor networks (WSNs) have gained a lot of attention recently due to the potential they provide for developing a plethora of cost-efficient applications. Although research on WSNs has been performed for more than a decade, only recently has the explosion of their potential applicability been identified. However, due to the fact that the wireless spectrum becomes congested in the unlicensed bands, there is a need for a next generation of WSNs, utilizing the advantages of cognitive radio (CR) technology for identifying and accessing the free spectrum bands. Thus, the next generation of wireless sensor networks is the cognitive wireless sensor networks (CWSNs). For the successful adoption of CWSNs, they have to be trustworthy and secure. Although the concept of CWSNs is quite new, a lot of work in the area of security and privacy has been done until now, and this work attempts to present an overview of the most important works for securing the CWSNs. Moreover, a discussion regarding open research issues is also given in the end of this work.

1. Introduction

WSNs are daily gaining more ground into our lives with applications ranging from construction monitoring and intelligent transport to smart home control and assisted living. Through the novel communication standards of the past decades such as Zigbee and IEEE 802.15.4, along with the pervasiveness of IEEE 802.11, the development of interoperability and commercial solutions has been enabled. Typically though, these solutions do suffer from strict deployment design and poor scalability. At the same time, the reliability of WSNs is a key topic for their mass adoption for more critical, rather than luxury or pilot, applications, such as the smart metering [1].

Cognitive radio (CR) features, such as the opportunistic spectrum (white space) usage, the introduction of secondary users in licensed bands, and the ability to learn the environment through sensing, present themselves as a mean to overcome spectrum shortage. Enabling such CR characteristics over "traditional" WSNs allows them to change their transmission parameters according to the radio environment and possibly enhance the reliability of WSNs in areas densely populated by wireless devices. These cognitive radio-imbued WSNs (CWSNs) can have access to new spectrum bands with better propagation characteristics. By adaptively changing system parameters like the modulation schemes, transmit power, carrier frequency, channel coding schemes, and constellation size, a wider variety of data rates can be achieved, especially when CWSNs operate on software-defined radios. This can improve device energy efficiency, network lifetime, and communication reliability.

The adoption of CR technology in CWSNs has largely improved network performance, but not without any cost. CWSNs, aside from being still open to a host of pure networking research issues, are also vulnerable to new types of threats. Attacks targeting a CWSN can come from both internal and external network sources. Adversaries can exploit vulnerabilities in different communication layers, many of which target the CR characteristics of the CWSN. There are also special types of attacks that try to infer sensitive information on the application and that execute in the sensors themselves [2]. Our work here aims to make a brief, yet succinct, overview of possible attacks on CWSNs. We therefore begin providing a background of WSNs and CWSNs in Sections 2 and 3, respectively. We then move to identify the common features and attacks in both of these types of networks in

Section 4. In Section 5, we specify attacks applicable only to CWSNs, and in Section 6 we detail security mechanisms for attack detection at different communication layers. Our work concludes with a discussion of open issues in Section 7.

2. Overview of Wireless Sensor Networks

WSNs have become widely available from the early 2000s, as sensing components and communication modules were already becoming cheap and small [3]. Monitoring the environment with such low cost devices became since then efficient, with a large volume of research having been conducted in the last almost two decades (one can trace the origins of WSNs in [4]). By now, WSN solutions are deployed in large scales and in various places and are being widely used in a variety of applications ranging from military [5] to agriculture [6] and from health care [7] to traffic management [8].

A WSN typically comprises a set of sensor nodes equipped with limited, low-power/short-range communication capabilities. Each of these nodes is a computational/communication platform which consists of (at least) a sensing module, a transceiver, a processor unit, and a power unit. The sensor node has typically small physical dimensions and its components are inexpensive. To make these sensor nodes more appealing, communication is commonly based on the license-free industrial, scientific, and medical (ISM) frequency band [9], in order to further limit operational costs for the overall WSN installation and to enable direct use of off-the-shelf communication solutions [10].

Depending on the application and deployment scenario, WSNs may vary in the communication paradigm they employ [11]. WSN applications set up to observe and consequently report to a "fusion center" the occurrence of an event (such as a fire), not needing to transmit continuously all measurements acquired by the sensors [12]. On the other hand, in scenarios such as pollution measurements [13] or seismic activity, the raw data can well be meaningful in its entirety; in such a case, the transmissions required would clearly be producing a heavy communication load; thus efficient channel access between the nodes as presented in [14] is required. These two extremely different cases indicate a mapping to the range of communication modes that may have to be used to handle the WSN most limiting resources: spectrum and energy (see [15, 16] and the references therein). A very rudimentary method to address these is WSN topological solutions which can be multihop [17], hierarchical [18], or one hop to infrastructure [19]. Each one in the respective references given has reasoning behind the underling spectrum management. Furthermore, in each of these cases a key factor that affects the system design is the power source and lifetime requirement of the WSN [20]. The node power unit, mentioned earlier, may be unlimited: for example, in indoor scenarios where the nodes can be directly plugged to the power grid. In such cases, energy plays little to no role. On the other hand, there can be extremely constrained scenarios such as the Smartdust, where literally every mWatt has to be accounted for, as the battery providing power is constrained even by its physical size, let alone its capacity. Energy harvesting [9] has recently been gathering significant attention as it can enable extension of the node lifetime, leveraging the environment resources (heat, motion, RF radiation, etc.).

3. Enhancing Wireless Sensor Networks with CR Technology

While the WSN solutions were progressing well into the late 2000s, the dramatically rising demand for wireless connectivity brought the spectrum utilization into the spotlight. Cognitive radio [21] and opportunistic communications, especially under the paradigms of opportunistic access or delay tolerant networking, came naturally into the frame of WSNs [22, 23]. Research into considering CR aspects for WSNs has thus begun [24, 25].

Opportunistic access is based on sending the transmissions over the "most suitable" spectrum band under a set of predefined application-driven requirements. With delay tolerance, a temporal aspect comes also into play: nodes can withhold data and transmit them at the "best" possible moment, subject to the application delay constraints. To enable these features, an additional process of dedicated spectrum sensing is required by the nodes, and in some cases local coordination can be used to enable the nodes to cooperatively infer about the radio spectrum usage at a specific area [26, 27]. This flexibility is further employed to adjust transmission parameters (modulation and coding schemes and transmission power) to reduce overall power consumption. Existing schemes developed to obtain spectrum awareness for cognitive radios in some cases consider the power consumption problem [28, 29], a clearly critical issue for CWSN. Reduced power consumption considered in CWSNs can not only extend the lifetime of sensor nodes but also limit the overall spectrum inefficiencies of the network, allowing for a substantial increase in spectrum utilization [30, 31].

4. Features and Common Attacks in WSNs and CWSNs

4.1. Common Features of WSNs and CWSNs. WSNs and CWSNs are two types of sensor networks that have a number of common characteristics. They consist of miniature devices, called motes or sensors that are severe resource constrained devices in terms of memory, processing, and energy [32, 33]. They usually do not perform any computation on the data they collect; they just forward this information to much more powerful devices (called sinks) for further processing.

The communication medium used for both WSNs and CWSNs has a broadcast nature and the used spectrum is split into several channels, depending on the protocol used. For example, there are up to 16 available channels for the IEEE 802.15.4 in the 2.4 GHz frequency band.

In both types of networks, the communication protocols used have a number of inefficiencies and vulnerabilities that allow potential attackers to launch a variety of destructive

attacks against these networks. The result of these attacks has catastrophic consequences including network performance deterioration, information theft, lifetime minimization, and battery depletion.

A multihop type of communication is often used in both types of networks (e.g., [34]) when data from a large and/or harsh area have to be sensed. Information flows from a sensor to a sink through multiple intermediate sensors that route packets according to an appropriate routing algorithm (e.g., RPL [35]). In a number of contributions, the network is split into several clusters and decisions are taken by the cluster heads in order to minimize sensors communication overhead and save energy, prolonging network's lifetime.

In both types of networks, network topology is highly dynamic and unpredictable without any central management. This is the case when sensors are deployed in harsh and volatile environments (e.g., [36, 37]). In such cases, adversaries can more easily attack and compromise the WSN.

4.2. Common Attacks against WSNs and CWSNs. The above common characteristics of WSNs and CWSNs make them vulnerable to a number of security threats. A diverse range of vulnerabilities are exploited by adversaries who can have several incentives, for example, network disruption, information theft, and so forth. In general, there are two types of attackers [38]: (i) external attackers that are not authorized participants of the sensor network and (ii) internal attackers that have compromised a legitimate sensor and use it to launch attacks in the network. Furthermore, attackers can be classified into passive and active. Passive attackers monitor network traffic without interfering with it. Their aim is to eavesdrop on the exchanged information and to acquire private data or to infer about information-sensitive applications that execute in the sensors (e.g., [2]). Active attackers disrupt network operation by launching several types of attacks that cause DoS (denial of service) in the WSN.

A severe DoS attack is jamming at the physical layer of the network. An adversary by creating interference, mainly through energy emission in the neighboring channels of the channel used by the sensor network [39], substantially increases the noise such that potential receivers become completely unavailable to receive and decode any information. This results in packet loss and further retransmissions by the senders that potentially lead to energy waste in the sensor network.

Jamming attacks can also be launched at the link layer. Here, an attacker can violate several characteristics of the communication protocol and cause packet collisions, exhausting sensors' resources. The authors in [40] show how a single adversary can cause severe performance degradation by violating several rules of the link layer protocol (back-off mechanism). Another popular attack is the Sybil attack where an adversary maliciously uses the identities of a number of sensors. This is achieved either by learning other sensors' identities or by fabricating new ones [41]. Furthermore, other types of attacks such as MAC spoofing [42] and ACK attacks [43] can cause confusion and packet loss in the network.

A major challenge in a WSN is maximizing its network lifetime by choosing the appropriate mode of communication. Single-hop communication, where the sensors communicate directly to a sink, is the flavour mode when the number of the sensors and the communication radius are small [44]. On the other hand, when the number of sensors is large (a typical case when a large area has to be covered by sensors) multihop communication is the most appropriate mode that saves sensors' energy, prolonging network's lifetime. In the multihop scenario, sensors have a dual role: they sense the environment and they also route the packets of their neighbors towards the sink (and vice versa). Packet forwarding and optimal path selection are performed by following an appropriate routing protocol. Adversaries can exploit several vulnerabilities and launch attacks against multihop sensor networks. Various attacks have been reported in the literature.

(i) *Selective Forwarding Attack.* Attackers drop the packets they have to route, randomly or selectively based on some rules (e.g., packets that originate from a specific sensor).

(ii) *Sinkhole Attack.* An attacker by broadcasting fake information makes the legitimate nodes believe that the attacker is attractive according to the routing protocol. If this attack is successful, neighboring sensors will forward their packets to the attacker that is then free to alter or steal information or drop the packets.

(iii) *Wormhole Attack.* This attack is performed by a number of colluding adversaries that forward packets between them through a direct long-distance and low-latency communication link (wormhole link). With this attack, legitimate sensors at a specific area of the network believe that they are close neighbors with sensors of another area that is however far away. This illusion creates confusion and affects routing within the network.

Except the above attacks that exploit several vulnerabilities in different layers of the communication stack, there is a special type of attack that aims to infer about information-sensitive application that executes in the sensors. Suppose that there is an on-body sensor network (e.g., [45]) consisting of a number of sensors that record high-sensitivity data such as the heart rate and oxygen saturation. Usually these applications transmit the sensed data to a sink in a periodic fashion [46]. Recent works [2, 46] have shown that adversaries can infer about these applications by passively monitoring the network traffic and detecting its periodic components that can finally reveal the potential medical applications. This becomes feasible using the appropriate signal processing techniques (e.g., the Lomb-Scargle periodogram) that discover traffic's periodic components even if it is encrypted.

5. Specific Attacks against CWSNs

As described in the previous section, WSNs and CWSNs have a number of common features and hence some common vulnerabilities that can be exploited by potential adversaries.

Nevertheless, CWSNs have two unique characteristics (that WSNs do not have) due to their cognitive nature [47].

(i) *Cognitive Capability.* It allows sensors to sense the environment for white spaces. Then, through a spectrum management process they decide upon which band to use for transmission and how to estimate the related-to-transmission physical layer parameters (frequency, modulation type, power, etc.). The cognitive cycle consists of several mechanisms: (i) radio environment, (ii) spectrum sensing, (iii) spectrum analysis, and (iv) spectrum decision.

(ii) *Reconfigurability.* It allows sensors to change on the fly their physical layer parameters and adapt to their environment. As sensors in CWSNs opportunistically use the fallow bands, they have to be flexible and vacate a band if a primary transmission is detected.

These unique characteristics make CWSNs vulnerable to a number of novel attacks. One of the most destructive attacks is called *primary user emulation attack* (PUEA). In this attack, an adversary mimics a primary user (PU) by transmitting fake incumbent signals [48]. Legitimate sensors will immediately evacuate the specific (under attack) frequency band, seeking for an alternative band to operate. Adversaries launching this attack can be of two types: (i) greedy sensors that emit the fake incumbent signals in order to make legitimate sensors evacuate the band in order to acquire its exclusive use and (ii) malicious sensors that aim to cause a DoS attack making sensors hop from band to band. Regardless of the type of the adversary, the PUEA attack can cause severe network disruption and a huge energy waste to the legitimate sensors. Figure 1 [47] shows that the PUEA attack affects all parts of the cognitive cycle.

As mentioned before, spectrum sensing is a fundamental operation and is one of the most challenging issues of the cognitive cycle. Spectrum sensing is the task of obtaining awareness about the spectrum usage and the possible presence of primary users [49]. During this operation, there is always the risk for the cognitive sensors not to correctly decode and hence detect the primary signals because of the shadow fading and hidden node effects. If this happens, harmful interference will be created to the primary transmitters. Collaborative spectrum sensing has been proposed as a solution to this problem [50]. In collaborative spectrum sensing, all sensors perform spectrum sensing and report their findings to a fusion centre (FC). The FC after performing a spectrum analysis procedure based on the sensors' reporting decides if a spectrum handoff has to be performed and at which frequency band. In a CWSN, the sink or the cluster heads (if the sensor network uses clusters) can have the role of the FC. However, if the network is not partitioned into clusters or the sink is far away from the majority of the sensors, this centralized scheme is not feasible. In such cases, distributed sensing can take place, where each sensor based on its own spectrum observation and the observations shared by its neighboring sensors makes its own spectrum decisions [51].

Adversaries can exploit the above mechanisms and affect FC's decision (or their neighbors' decision in distributed sensing) by sending false observations regarding spectrum usage. This attack is called *spectrum sensing data falsification* attack (SSDF). SSDF attackers can report that a specific band is vacant when it is not or that is occupied by primary signals when it is not. In the first case, harmful interference to the primary users will be created, while in the latter legitimate sensors will keep performing costly (in terms of energy) spectrum handoffs. Attackers can have different motives: (i) they can be greedy users that continuously report that a specific band is occupied in order to acquire its exclusive use and (ii) they can be malicious nodes that, by sending false observations, aim to create interference to primary transmitters or create a DoS attack on the network due to the continuous spectrum handoff of the legitimate sensors. SSDF attacks can also be initiated by unintentionally misbehaving sensors that report false observations because some parts of their software or hardware components are malfunctioning. This type of attack can substantially degrade network's performance as the authors in [52] show. Regarding the cognitive cycle, the SSDF attack affects the spectrum analysis and spectrum decision operations (Figure 1).

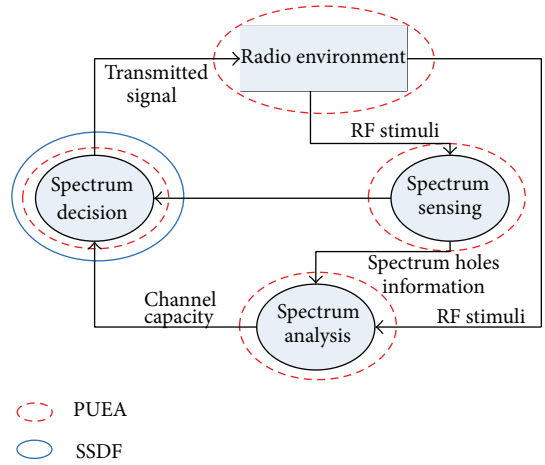


Figure 1: The cognitive cycle [47].

6. Security, Privacy, and Reliability Mechanisms for CWSNs

6.1. Security. Securing a WSN is of paramount importance, and for this reason a large number of contributions exist in the literature for the detection and mitigation of attacks against this type of networks. Depending on the attack type, different strategies and algorithms are followed.

6.1.1. Physical Layer Attack Detection. As mentioned in Section 4.2, jamming at the physical layer can cause disruptive DoS attacks in a WSN. The detection techniques try to (almost) instantly detect that a jamming attack is taking place by considering various metrics. The authors in [53] use the signal-to-interference-plus-noise ratio (SINR) as the metric that can signal the jamming attack. The recorded SINR values are fed to a cumulative-sum algorithm that

is able to detect abrupt changes that are caused by the attacker's presence. The performance of this anomaly-based detection algorithm is augmented if several monitors are used in a collaborative intrusion detection scheme. In [54], the definition of several types of attackers is given, and jamming detection is performed by using multiple if-else statements considering as metrics the *packet delivery ratio*, the *bad packet ratio*, and the *energy consumption amount*. In [55], a distributed anomaly detection algorithm is presented based on simple thresholds and a method for combining measurements using Pearson's product moment correlation coefficient. RF jamming attacks are the focus of [56] where the proposed algorithm applies high order crossings, a spectral dissemination technique that distinguishes normal scenarios from two types of defined attackers. The detection algorithm is based on thresholds considering the signal strength and location information. The authors in [57] propose DEEJAM, a defensive mechanism that uses an IEEE 802.15.4-based hardware. Here, the proposed algorithm hides messages from a jammer, evades its search, and reduces the impact of the corrupted messages.

6.1.2. Link Layer Attack Detection. Contributions that study the detection of attacks at the link layer include [40]. Here, an anomaly-based algorithm is presented considering the ratio of the corrupted packets over the correctly decoded packets as the metric that reveals jamming when the attacker's energy is emitted on the same channel. In [58], the authors explore energy-efficient attacks targeting three WSN protocols: (i) S-MAC, (ii) B-MAC, and (iii) L-MAC. As a countermeasure they suggest the use of shorter data packets for the L-MAC and high duty cycle for the S-MAC. Link layer misbehaviour in [59] is detected by applying a nonparametric cumulative-sum algorithm considering the expected back-off value of the honest participants. MAC address spoofing detection in WSN is studied in [60]. In that work, an approach based on the Gaussian mixture models that considers RSS (received-signal-strength) profiles is used to detect if a MAC address is spoofed. RSS is a metric that is hard to forge arbitrarily, and it highly depends on the transmitter's location. The authors in [42] propose an algorithm that leverages the sequence number field carried by the data packets. This algorithm records the sequence number of each received frame and that of the last frame coming from the same source node. When the gap between the current sequence number and the last recorded one is within a specific range, it is considered as abnormal. For each abnormal frame, a verification process follows to declare the specific frame as normal or spoofed.

Regarding the Sybil attack detection, the algorithm in [61] uses the ratios of the RSSI (received-signal-strength indicator) recorded in a number of sensor monitors when a packet is transmitted within their communication range. If these ratios are very close to the ratios computed when a packet with a different identity is used, the corresponding transmitter is flagged as a Sybil attacker. In [62], the detection algorithm exploits the characteristic that every Sybil (forged) sensor has the same set of neighbors as they are created by the same adversary. It detects the Sybil attack by comparing the information collected from neighboring sensors (contained in small messages). In [63], Sybil attacks are detected by exploiting the spatial variability of radio channels in environments with rich scattering. An enhanced physical layer authentication scheme is used for both wideband and narrowband wireless systems.

6.1.3. Network Layer Attack Detection. As described in Section 4.2, a large number of vulnerabilities of the routing protocols can be exploited in sensor networks. Different countermeasures have been proposed for the detection of these attacks. In [64], a lightweight scheme uses a multihop acknowledgment technique to launch alarms when responses from intermediate sensors are missing. Each time a sensor receives a data packet, it sends an ACK to the sensor that handled the packet in the previous hop. If a sensor receives less than a number of ACK packets within a specified time, it suspects that the previous report it forwarded has been dropped by a malicious sensor. If this is the case, it sends an alarm packet to the sink, reporting its next-hop sensor as a potential malicious sensor. The sink after it receives all alarm packets infers about the malicious sensors. The authors in [65] propose a centralized scheme with the use of support vector machines (SVMs). A 2D SVM is initially trained when no attacker is present, using the hop count and the measured bandwidth at the sink as features. At run time, the detection algorithm based on the SVM executes at the sink. A different approach is followed in [66] where each sensor observes the behavior of its neighbors recording the number of packets they forward, along with the source address of the originating sensor. Based on these observations, it updates a trust metric for each of its neighbors that reveals the potential attackers. After a sensor has been labelled as an attacker, the routing tables are modified in order to isolate that sensor from the network.

For the detection of the sinkhole attacks, a distributed detection scheme is presented in [67]. Every sensor S_i is set in promiscuous mode and records the route update packets transmitted by its neighbors. Furthermore, two rules have been defined that if violated, an alert message is broadcasted: (i) if sender's ID matches S_i's ID and (ii) if sender's ID does not belong to the known IDs of S_i's neighbors. This detection scheme also employs a collaborative detection algorithm that reveals the potential attacker based on an intersection computation of the information carried by the alert messages. Ngai et al. [68] propose a detection algorithm that consists of two steps: (i) it locates a list of suspected sensors by checking data consistency based on the information sensors report to the sink and (ii) it labels a sensor as an attacker by analyzing the network flow information (represented by directed edges between communicating sensors). The authors in [69] show that shortest-path routing protocols select a series of paths whose length exhibits a log-normal distribution. Based on this observation, they propose an anomaly detection algorithm by deriving tolerance limits from the log-normal distribution of path lengths when no attacker is present.

Regarding the wormhole detection, the scheme proposed in [70] considers the round-trip time (RTT) between an originating sensor and its destination. RTT depends on how far the intermediate sensors are located. If a wormhole attack is in progress, RTT can significantly increase, as packets are replicated in a different part of the network from colluding attackers. In [71], a localized scheme based on connectivity graphs is proposed. It seeks for *forbidden substructures* in the connectivity graphs that should not be present under normal circumstances. The authors in [72] propose a distributed detection algorithm that detects wormhole attacks based on the distortions these attacks create in the network. This scheme uses a hop counting technique as a probe procedure, reconstructing local maps for each sensor, and then a diameter feature that depends on the number of neighboring nodes, for anomaly detection.

6.1.4. Detection of Attacks That Exploit Vulnerabilities of the Cognitive Nature of CWSNs. A possible framework for securing cognitive radio networks has been proposed in [73] and can easily be extended to secure CWSNs. This framework attempts to identify the mechanisms that can mitigate the specific attacks on cognitive radio networks. As discussed in Section 5, there are two major types of attacks that can be launched against CWSNs: (i) PUEAs and (ii) SSDF attacks. Regarding the detection of the PUEAs, there are a large number of significant contributions that split into two main categories: (i) location-based and (ii) non-location-based contributions. Location-based contributions assume that the locations of the primary transmitters are known a priori.

The work in [48] considers both the location information of the primary transmitter and the RSS values collected by a separate sensor network each time a primary transmission is taking place. Based on the RSS measurements the location of the transmitter is estimated, and if it is different than the (already) known location of the legitimate primary transmitter, an alarm is triggered. Jin et al. [74] developed an algorithm that considers the received power measured at the radio interfaces of the secondary users (SUs) in a specific band. Then, by using Fenton's approximation and Wald's sequential probability radio test, they decide on the corresponding hypothesis about the presence or not of a PUEA attacker. The received power is also considered in [75] where the authors propose a variance method to detect the attack. This scheme first estimates the variance of the received power from the primary transmitter, and then it determines whether a received signal is from the primary transmitter or from an attacker.

In non-location-based algorithms like in [76], the locations of the primary transmitters are not required to be known. The authors state that the channel impulse response can be revealed if a primary transmitter has moved to a different location. Their approach uses a *helper node* (HN) that is located very close to a primary transmitter in a fixed location. This node is used as a bridge between the SUs and the primary transmitter by allowing SUs to verify cryptographic signatures by HN's signals and then obtain HN's link signals in order to verify primary transmitter's signals. The authors show that, by using the first and second multipath components measured at HN, they can verify if the transmitted signal belongs to the legitimate primary user or it is fake. The scheme presented in [77] uses a public key cryptography mechanism where a primary transmitter integrates its transmitted data with cryptographic signatures. Each SU that detects a primary signal attempts to verify its integrated signatures. If verification fails, the signal is characterized as fake.

Regarding the SSDF attack detection, in [52] a centralized algorithm calculates the trust values of SUs based on their past record. Additionally, consistency checks are performed because the trust values can become unstable if an attacker is present or there is not enough information. If the consistency value and the trust value of an SU drop below a specific threshold, the specific SU is characterized as an attacker. Rawat et al. [78] propose a centralized scheme that computes a reputation metric for each SU based on SU's past observation, and the decision is made by the FC during that round of observations. If there is a decision mismatch, SU's reputation metric is increased by one, and if it becomes larger than a predefined threshold, SU is labelled as an attacker. Reputation metrics are also used by other similar contributions like in [79, 80].

6.2. Privacy. Although security attacks in WSNs have been very extensively researched until now, "privacy" attacks are a not so common research topic. Most works until now have focused mainly on protecting the location privacy of the sensor nodes, while others focus on protecting the traffic of the data that are transmitted by the nodes. However, when sensors are enhanced with CR technology, the traditional WSN privacy attacks still exist, with the addition of other attacks for eavesdropping on the sensing data (in collaborative spectrum sensing) and the context of the exchanged sensor data, for impersonating the PU and against the anonymity of a sensor node. In this section the common attacks against privacy on CWSN are described, together with the existing mechanisms for mitigating these attacks.

6.2.1. CR Location Privacy. Location privacy is a major research topic in cognitive WSNs due to the fact that the spectrum opportunities (namely, the unoccupied spectrum frequencies or the white spaces) are heavily depending on the location of both the sensor nodes and the PUs. The received PU signal at the sensor nodes is highly related to the distance between the sensor nodes and a malicious user can identify the sensor node location using geolocation mechanisms. Furthermore, in participatory sensing [81] the data from the sensor nodes are usually tagged with location and the time.

According to [82], the respective location privacy attacks can be either external (combined with eavesdropping) or internal. An external attacker can intercept the spectrum sensing reports that are exchanged throughout the CWSN by eavesdropping on the communication of the sensor nodes either with each other or with the FC (in case of a centralized spectrum sensing system). That way, the attacker is able to

know the received PU signals of all sensor nodes and by correlating the data with its own sensing reports, he is able to identify the location of the sensor nodes. An internal attacker can be either another node participating in the collaborative sensing or the fusion center (or an attacker impersonating the fusion center). That way the attacker seems to be a legitimate node that receives the sensing reports from all other nodes and can easily compromise their location by correlating the data with his physical location. An internal attacker can also exploit the results of the aggregated sensing reports that are being transmitted by the FC. That way, comparing the reports before and after the inclusion of a new node in the network, it is easy to identify its location.

Mitigation. For preserving the privacy of cognitive sensor nodes, in [82] a combination of techniques for cryptography and sensing data randomization has been proposed. The first technique uses the concept of secrets [83] and each sensor encrypts its sensing data in such a way that the FC should get all reports in order to be able to decrypt the aggregated sensing report. That way, when a malicious user intercepts the reports from a specific sensor or from all sensors in an area, he will not be able to decrypt these reports, hence the sensors' locations cannot be estimated.

Another proposal [82] for protecting the location of cognitive sensors includes the transmission of dummy sensing reports from one of the legitimate nodes or the fusion center when a new node is joining or leaving the network. Although this can degrade the performance of collaborative sensing, an appropriate selection of the dummy report and its weight on the overall sensing aggregation can have a minimal impact, without affecting significantly the sensing result.

Proposals for ensuring location privacy in participatory sensing include the anonymization of sensing reports using the principle of k-anonymity [84–87], which assumes that at least k users are located at the same area, and thus they tag their sensing reports with an area "ID" and not with their actual location information. That way, if an attacker eavesdrops on the reports of the sensor nodes, only an abstract view of the general area of the users could be extracted and not an actual location. However, the performance of such a sensing system is heavily depending on the size of the area, because a small area can result in an optimum sensing result but can also give enough information to the attacker to identify the location of the sensor nodes. On the other hand, a large area may preserve the nodes' location information but can degrade significantly the performance of the participatory sensing system.

6.2.2. Sensed Data Privacy. Like traditional WSNs, CWSNs are deployed for getting automated measurements and transmitting them to an application server for processing. This information may be sensitive in some applications and must be protected from unauthorized access and use. For example, hijacking the information sent by sensors measuring the energy consumption of devices in a household may reveal the presence/absence of the habitants, which could be utilized by burglars. Respective attacks against the sensor data include eavesdropping, impersonation, and traffic analysis [88].

Eavesdropping (or passive monitoring) is a very common attack on WSNs, under which an attacker is listening to the communication channel of the sensor nodes and intercepts their data. In this attack, the malicious node is hidden from the sensor nodes because it does not communicate directly with them. Under the impersonation attack, the malicious node impersonates either a legitimate node or the FC and gets the data directly from the legitimate sensor nodes. This attack can be the first point to launch other attacks changing the data and transmitting false data to the other nodes. The traffic analysis is used by attackers to extract the context of data that are transferred by the sensors and is achieved by analysing the traffic patterns from eavesdropping on the wireless links. Using the traffic analysis attack, a malicious node can also identify some nodes that have a special role in the CWSN (i.e., who has the role of the FC).

Mitigation. Targeting avoidance of the disclosure of the sensed data to unauthorized recipients, several proposals have been made in the literature, which mainly focus on anonymity schemes or on information flooding. Using anonymization, the data sent by a legitimate node do not contain personal information that can be used to track back the measurements to the originating sensor node [89]. In [90], a framework for context-aware privacy of sensor data is proposed, which includes a two-step process of (i) identifying which data will be shared and (ii) obfuscating the data before transmitting them. Although most previous anonymization proposals were focused on protecting sensor location information [82, 91], they can be relatively easily adapted to the sensed data that the nodes are transmitting. Information flooding is another technique that can be used to protect the data privacy in CWSNs, as proposed in [92], which discusses the fact that probabilistic flooding can give good protection to the node information while being energy efficient.

7. Conclusion

WSNs and CWSNs are two similar sensor network types with quite a few common features. Recently there has been an explosion of Smart City applications for providing advanced ICT-based services to citizens with the use of enhanced WSN networks. For the realization of such applications a plethora of sensing and actuating devices are usually installed either in a city area or within buildings. In this context, the WSNs will be playing a significant role in the everyday life of people, and thus their security is of great importance. This explosion in the number of wireless sensing and actuating devices in city areas together with the continuous installation of many (public and private) wireless access networks in these areas has resulted in congestion in the unlicensed spectrum bands (ISM bands around 2.4 GHz) that are used for both WSNs and Wi-Fi. For mitigating the congestion effects on the WSN networks, there are proposals to equip the latter with CR technology forming the CWSNs, which on the one hand

Table 1: Attacks against CWSNs.

Type of attack	OSI layer	Characteristic	Common with WSN
Jamming	Physical layer	DoS attack creating interference, increasing packet loss and collisions	Yes
Back-off attack	Link layer	An attacker causes severe performance degradation by minimizing the CWmin and thus his back-off period	Yes
Sybil attack	Cross layer	Stealing sensors identities, that is, MAC address, IP address, and so forth	Yes
MAC spoofing	Link layer	Alternating a MAC address on a network interface can help an unauthorized intruder enter a secure network	Yes
Selective forwarding attack	Network layer	Attackers drop packets they have to route	Yes
Sinkhole attack	Network layer	Attacker broadcasts false routing related information so that neighbouring nodes send them their packets and steals information or drops them	Yes
Wormhole attack	Network layer	Adversaries exchange packets through a long-distance and low-latency links affecting routing making legitimate sensors believe that they are neighbours with sensors of another area	Yes
PUEA	Physical layer	Adversaries mimic PU so that hey exploit unused frequencies that the other nodes assume as occupied by the PU	No
SSDF attack	Physical layer	Attackers provide false information regarding spectrum occupancy	No
Location privacy attacks	Physical layer	Attackers intercept signals and sensing reports so that with data correlation they can identify the sensor location	No
Sensed data privacy attacks	Physical/link layer	Attackers eavesdrop the channel and analyse the traffic to intercept the sensed data that are transmitted by the sensors	Yes

solves several issues of traditional WSNs security-wise but on the other introduces new security threats.

Securing WSNs and CWSNs is of key importance, and a large pool of contributions from the literature for the detection and mitigation of attacks against these networks has been presented in this paper. Furthermore, an overview of the most common attacks against CWSN is presented in Table 1. Depending on the attack type, different strategies and algorithms are followed. Exploiting the CR features of CWSN enables two major classes of attacks that can be launched against them: (i) PUEAs and (ii) SSDF attacks. Regarding the detection of the PUEAs a significant number of contributions exist which can be broken into two categories: (i) location-based and (ii) non-location-based contributions. For the former the key challenge is the detection of the attacker's location, an issue that is open in many other problems of wireless networking. The SSDF attack detection in the literature presented here is primarily based on the notions of reputation and trust, given the collaborative nature of the proposed solutions. Regarding privacy, the most common attacks are those against identifying the location of the cognitive sensors node and those against intercepting the sensing data.

Although much research has been done in the literature regarding the security of the CWSNs, there are still several challenges and open research issues remaining. One of the most important challenges is related to introducing trust within the CWSN architecture. Although several attempts for mitigating SSDF attacks are introducing reputation mechanisms for the cognitive nodes, these can be considered as an "add-on" feature, while a trust framework embedded within the cognitive nodes not only addresses the SSDF attacks but also ensures the complete trustworthy operation (starting from the sensed data and going all the way up to ensuring the trustworthiness of the applications that run on the nodes) of the cognitive nodes. Another open challenge is related to designing lightweight cryptographic algorithms that could run on the very resource-limited cognitive sensor nodes, focusing on private-key cryptography, efficient key distribution schemes for symmetric key cryptography, and efficient key management protocols for public key cryptography. Regarding routing, in CWSNs there is a need for further research on secure routing schemes taking into account the spectrum assigned to each one of the intermediate nodes, as well as the mobility of the nodes and the potential scalability and efficiency issues. Moreover, in data aggregation mechanisms there is a need for further research on enhancing the data aggregation and securing it against malicious cognitive users, introducing trust and security metrics. Other open research issues regarding security in CWSNs that need to be addressed in future research include the use of geolocation information for improving security, that is, in PUEA attacks, the investigation of intelligent physical layer security mechanisms that exploit CR characteristics, the development of distributed mechanisms against SSDF attacks, and the design of efficient cooperative mechanisms against malicious nodes and intruders.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreements nos. 609094 and 612361.

References

[1] A. Liotta, D. Geelen, G. Kempen, and F. Hoogstraten, "A survey on networks for smart-metering systems," *International Journal of Pervasive Computing and Communications*, vol. 8, pp. 23–52, 2012.

[2] A. Fragkiadakis and I. Askoxylakis, "Malicious traffic analysis in wireless sensor networks using advanced signal processing techniques," in *Proceedings of the 14th International Symposium and Workshops on aWorld of Wireless, Mobile and Multimedia Networks (WoWMoM '13)*, pp. 1–6, 2013.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[4] S. DUST, "SMART DUST Autonomous sensing and communication in a cubic millimeter," http://robotics.eecs.berkeley.edu/~pister/SmartDust/.

[5] M. Winkler, M. Street, K. D. Tuchs, and K. Wrona, "Wireless sensor networks for military purposes," in *Autonomous Sensor Networks*, vol. 13 of *Springer Series on Chemical Sensors and Biosensors*, pp. 365–394, 2013.

[6] Y. Xiaoqing, W. Pute, W. Hana, and Z. Zhanga, "A survey on wireless sensor network infrastructure for agriculture," *Computer Standards and Interfaces*, vol. 35, no. 1, pp. 59–64, 2013.

[7] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1947–1960, 2010.

[8] A. Pascale, M. Nicoli, F. Deflorio, B. Dalla Chiara, and U. Spagnolini, "Wireless sensor networks for traffic management and road safety," *IET Intelligent Transport Systems*, vol. 6, no. 1, pp. 67–77, 2012.

[9] E. A. Fan Zhang, "A batteryless 19uw mics/ism-band energy harvesting body area sensor node soc," in *Proceedings of the Solid-State Circuits Conference Digest of Technical Papers (ISSCC '12)*, 2012.

[10] B. Buchli, F. Sutton, and J. Beutel, "Gps-equipped wireless sensor network node for high-accuracy positioning applications," in *Wireless Sensor Networks*, vol. 7158 of *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2012.

[11] P. Santi, "Topology control in wireless ad hoc and sensor networks," *ACM Computing Surveys*, vol. 37, no. 2, pp. 164–194, 2005.

[12] G. Wittenburg, "Cooperative event detection in wireless sensor networks," *Communications Magazine*, vol. 50, no. 12, 2012.

[13] R. P. Khedo and A. Mungur, "A wireless sensor network air pollution monitoring system," *International Journal of Wireless and Mobile Networks*, vol. 2, no. 2, 2012.

[14] A. Antonopoulos and C. Verikoukis, "Network-coding-based cooperative ARQ medium access control protocol for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 601321, 2012.

[15] J. A. S. G. Zhou and S. H. Son, "Crowded spectrum in wireless sensor networks," in *Proceedings of the 3rd Workshop on Embedded Networked Sensors*, 2006.

[16] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: a survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009.

[17] T. F. Abdelzaher, S. Prabh, and R. Kiran, "On real-time capacity limits of multihop wireless sensor networks," in *Proceedings of the 25th IEEE International Real-Time Systems Symposium (RTSS '04)*, pp. 359–370, December 2004.

[18] K. Iwanicki and M. Van Steen, "On hierarchical routing in wireless sensor networks," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN '09)*, pp. 133–144, April 2009.

[19] J. Zhang, L. Shan, H. Hu, and Y. Yang, "Mobile cellular networks and wireless sensor networks: toward convergence," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 164–169, 2012.

[20] I. F. Akyildiz, W. Lee, and K. R. Chowdhury, "CRAHNs: cognitive radio ad hoc networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 810–836, 2009.

[21] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[22] S. Huang, X. Liu, and Z. Ding, "Opportunistic spectrum access in cognitive radio networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications*, pp. 2101–2109, April 2008.

[23] O. S. A. Lindgren and A. Doria, "Probabilistic routing in intermittently connected networks," *SIGMOBILE Mobile Computing Communications Review*, vol. 23, no. 2, 2003.

[24] I. F. Akyildiz, W. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[25] O. B. Akan, O. B. Karli, and O. Ergul, "Cognitive radio sensor networks," *IEEE Network*, vol. 23, no. 4, pp. 34–40, 2009.

[26] S. Maleki, A. Pandharipande, and G. Leus, "Energy-efficient distributed spectrum sensing for cognitive sensor networks," *IEEE Sensors Journal*, vol. 11, no. 3, pp. 565–573, 2011.

[27] E. Tragos, S. Zeadally, A. Fragkiadakis, and V. Siris, "Spectrum assignment in cognitive radio networks: a comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1108–1135, 2013.

[28] J. A. Han, W. S. Jeon, and D. G. Jeong, "Energy-efficient channel management scheme for cognitive radio sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1905–1910, 2011.

[29] Y. Xu, C. Wu, C. He, and L. Jiang, "A cluster-based energy efficient mac protocol for multi-hop cognitive radio sensor networks," in *Proceedings of the Global Communications Conference (GLOBECOM '12)*, 2012.

[30] A. S. Zahmati and X. Fernando, "Application-specific spectrum sensing method for cognitive sensor networks," *IET Wireless Sensor Systems*, vol. 3, no. 3, pp. 193–204, 2013.

[31] E. Tragos and V. Angelakis, "Cognitive radio inspired m2m communications," in *Proceedings of the 16th International Symposium on Wireless Personal Multimedia Communications (WPMC '13)*, pp. 1–5, June 2013.

[32] B. Otal, C. Verikoukis, and L. Alonso, "Efficient power management based on a distributed queuing MAC for wireless sensor networks," in *Proceedings of the IEEE 65th Vehicular Technology Conference (VTC '07)*, pp. 105–109, April 2007.

[33] J. Alonso-Zarate, E. Stavrou, A. Stamou, P. Angelidis, L. Alonso, and C. Verikoukis, "Energy-efficiency evaluation of a medium access control protocol for cooperative ARQ," in *Proceedings of the IEEE International Conference on Communications (ICC '11)*, June 2011.

[34] N. Gazoni, V. Angelakis, V. A. Siris, and B. Raffaele, "A framework for opportunistic routing in multi-hop wireless networks," in *Proceedings of the 7th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN '10)*, pp. 50–57, October 2010.

[35] U. Herberg and T. Clausen, "A comparative performance study of the routing protocols LOAD and RPL with bi-directional traffic in low-power and lossy networks (LLN)," in *Proceedings of the 8th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN '11)*, pp. 73–80, November 2011.

[36] G. Werner-Allen, K. Lorincz, M. Welsh et al., "Deploying a wireless sensor network on an active volcano," *IEEE Internet Computing*, vol. 10, no. 2, pp. 18–25, 2006.

[37] V. Bychkovsky, K. Chen, M. Goraczko et al., "The CarTel mobile sensor computing system," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 383–384, November 2006.

[38] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 2, pp. 52–73, 2009.

[39] A. G. Fragkiadakis, V. A. Siris, and A. P. Traganitis, "Effective and robust detection of jamming attacks," in *Proceedings of the Future Network and Mobile Summit*, June 2010.

[40] A. Fragkiadakis, E. Tragos, T. Tryfonas, and I. Askoxylakis, "Design and performance evaluation of a lightweight wireless early warning intrusion detection prototype," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 37, pp. 1–18, 2012.

[41] C. E. Loo, M. Y. Ng, C. Leckie, and M. Palaniswami, "Intrusion detection for routing attacks in sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2, no. 4, pp. 313–332, 2006.

[42] F. Guo and T. Chiueh, "Sequence number-based mac address spoof detection," in *Proceedings of the 8th international conference on Recent Advances in Intrusion Detection (RAID '05)*, pp. 1–20, 2005.

[43] Y. Xiao, S. Sethi, H. Chen, and B. Sun, "Security services and enhancements in the IEEE 802.15.4 wireless sensor networks," in *Proceedings of the IEEE Global Telecommunications Conference(GLOBECOM '05)*, pp. 1796–1800, December 2005.

[44] J. F. Shi, X. X. Zhong, and S. Chen, "Study on communication mode of wireless sensor networks based on effective result," *Journal of Physics: Conference Series*, vol. 48, no. 1, article 245, pp. 1317–1321, 2006.

[45] B. Otal, L. Alonso, and C. Verikoukis, "Highly reliable energy-saving mac for wireless body sensor networks in healthcare systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 553–565, 2009.

[46] L. Buttyan and T. Holczerr, "Traffic analysis attacks and countermeasures in wireless body area sensor networks," in *proceedings of the International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMom '12)*, pp. 1–6, 2012.

[47] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Communications Surveys and Tutorials*, vol. 15, pp. 428–445, 2013.

[48] R. Chen, J. Park, and J. H. Reed, "Defense against primary user emulation attacks in Cognitive Radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 25–37, 2008.

[49] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[50] S. Zarrin and T. J. Lim, "Cooperative quickest spectrum sensing in cognitive radios with unknown parameters," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, December 2009.

[51] Z. Tian, E. Blasch, W. Li, G. Chen, and X. Li, "Performance evaluation of distributed compressed wideband sensing for cognitive radio networks," in *Proceedings of the 11th International Conference on Information Fusion (FUSION '08)*, July 2008.

[52] W. Wang, H. Li, Y. Sun, and Z. Han, "Attack-proof collaborative spectrum sensing in cognitive radio networks," in *Proceedings of the 43rd Annual Conference on Information Sciences and Systems (CISS '09)*, pp. 130–134, March 2009.

[53] A. Fragkiadakis, V. Siris, N. Petroulakis, and A. Traganitis, "Anomaly-based intrusion detection of jamming attacks, local versus collaborative detection," *Wireless Communications and Mobile Computing*, 2013.

[54] M. Cakiroglou and T. Ozcerit, "Jamming detection mechanisms for wireless sensor networks," Proceedings of the 3rd International Conference on Scalable Information Systems, 2008.

[55] A. Sheth, C. Doerr, D. Grunwald, R. Han, and D. Sicker, "MOJO: a distributed physical anomaly detection system for 802.11 WLANs," in *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services*, pp. 191–204, June 2006.

[56] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 46–57, May 2005.

[57] A. D. Wood, J. A. Stankovic, and G. Zhou, "DEEJAM: defeating energy-efficient jamming in IEEE 802.15.4-based wireless networks," in *Proceedings of the 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '07)*, pp. 60–69, June 2007.

[58] Y. W. Law, M. Palaniswami, L. V. Hoesel, J. Doumen, P. Hartel, and P. Havinga, "Energy-efficient link-layer jamming attacks against wireless sensor network MAC protocols," *ACM Transactions on Sensor Networks*, vol. 5, no. 1, article 6, 2009.

[59] A. A. Cárdenas, S. Radosavac, and J. S. Baras, "Evaluation of detection algorithms for MAC layer misbehavior: theory and experiments," *IEEE/ACM Transactions on Networking*, vol. 17, no. 2, pp. 605–617, 2009.

[60] Y. Sheng, K. Tan, G. Chen, D. Kotz, and A. Campbell, "Detecting 802.11 MAC layer spoofing using received signal strength," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 2441–2449, April 2008.

[61] M. Demirbas and Y. Song, "An RSSI-based scheme for sybil attack detection in wireless sensor networks," in *Proceedings of the International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '06)*, pp. 564–568, June 2006.

[62] K. Ssu, W. Wang, and W. Chang, "Detecting Sybil attacks in Wireless Sensor Networks using neighboring information," *Computer Networks*, vol. 53, no. 18, pp. 3042–3056, 2009.

[63] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based detection of sybil attacks in wireless networks," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 492–503, 2009.

[64] B. Yu and B. Xiao, "Detecting selective forwarding attacks in wireless sensor networks," in *Proceedings of the 20th International Parallel and Distributed Processing Symposium (IPDPS '06)*, pp. 1–8, 2006.

[65] S. Kaplantzis, A. Shilton, N. Mani, and Y. A. Şekerciğlu, "Detecting selective forwarding attacks in wireless sensor networks using support vector machines," in *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP '07)*, pp. 335–340, December 2007.

[66] Y. Cho and G. Qu, "Detection and prevention of selective forwarding-based denial-of-service attacks in wsns," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 205920, 16 pages, 2013.

[67] I. Krontiris, T. Dimitriou, T. Giannetsos, and M. Mpasoukos, "Intrusion detection of sinkhole attacks in wireless sensor networks," in *Algorithmic Aspects of Wireless Sensor Networks*, pp. 150–161, 2008.

[68] E. C. H. Ngai, J. Liu, and M. R. Lyu, "An efficient intruder detection algorithm against sinkhole attacks in wireless sensor networks," *Computer Communications*, vol. 30, no. 11-12, pp. 2353–2364, 2007.

[69] D. Dallas, C. Leckie, and K. Ramamohanarao, "Hop-count monitoring: detecting sinkhole attacks in wireless sensor networks," in *Proceedings of the 15th IEEE International Conference on Networks (ICON '07)*, pp. 176–181, November 2007.

[70] Z. Tun and A. Maw, "Wormhole attack detection in wireless sensor networks," in *Proceedings of the World Academy of Science, Engineering and Technology*, pp. 545–550, 2008.

[71] R. Maheshwari, J. Gao, and S. R. Das, "Detecting wormhole attacks in wireless networks using connectivity information," in *Proceedings of the 26th IEEE International Conference on Computer Communications*, pp. 107–115, May 2007.

[72] Y. Xu, G. Chen, J. Ford, and F. Makedon, "Detecting wormhole attacks in wireless sensor networks," *IFIP International Federation for Information Processing*, vol. 253, pp. 267–279, 2007.

[73] A. Mihovska, R. Prasad, E. Tragos, and V. Angelakis, "Design considerations for a cognitive radio trust and security framework," in *Proceedings of the IEEE 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD '12)*, pp. 156–158, September 2012.

[74] Z. Jin, S. Anand, and K. P. Subbalakshmi, "Detecting primary user emulation attacks in dynamic spectrum access networks," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, June 2009.

[75] Z. Chen, T. Cooklev, C. Chen, and C. Pomalaza-Ráez, "Modeling primary user emulation attacks and defenses in cognitive radio networks," in *Proceedings of the 28th International Performance Computing and Communications Conference (IPCCC '09)*, pp. 208–215, December 2009.

[76] Y. Liu, P. Ning, and H. Dai, "Authenticating primary users' signals in cognitive radio networks via integrated cryptographic and wireless link signatures," in *Proceedings of the 31st IEEE Symposium on Security and Privacy (SP '10)*, pp. 286–301, May 2010.

[77] C. N. Mathur and K. P. Subbalakshmi, "Digital signatures for centralized DSA networks," in *Proceedings of the 4th Annual IEEE Consumer Communications and Networking Conference (CCNC '07)*, pp. 1037–1041, January 2007.

[78] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Countering Byzantine attacks in cognitive radio networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 3098–3101, March 2010.

[79] R. Chen, J. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in *Proceedings of the 27th Conference on Computer Communications (INFOCOM '08)*, pp. 1876–1884, 2008.

[80] E. Noon and H. Li, "Defending against hit-and-run attackers in collaborative spectrum sensing of cognitive radio networks: a point system," in *Proceedings of the 71st Vehicular Technology Conference (VTC '10)*, May 2010.

[81] J. Burke, D. Estrin, M. Hansen et al., "Participatory sensing," in *Proceedings of the Workshop on World-Sensor-Web (WSW '06), 2006*, pp. 117–134.

[82] Z. Gao, H. Zhu, S. Li, S. Du, and X. Li, "Security and privacy of collaborative spectrum sensing in cognitive radio networks," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 106–112, 2012.

[83] E. Shi, R. Chow, T. H. Hubert Chan, D. Song, and E. Rieffel, "Privacy-preserving aggregation of time-series data," in *Proceedings of the NDSS Symposium*, 2011.

[84] K. L. Huang, S. S. Kanhere, and W. Hu, "Towards privacy-sensitive participatory sensing," in *Proceedings of the 7th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom '09)*, March 2009.

[85] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "AnonySense: privacy-aware people-centric sensing," in *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, pp. 211–224, ACM, New York, NY, USA, June 2008.

[86] A. Kapadia, N. Triandopoulos, C. Cornelius, D. Peebles, and D. Kotz, "Anonysense: opportunistic and privacy-preserving context collection," in *Proceedings of the 6th International Conference on Pervasive Computing*, pp. 280–297, Springer, Berlin, Germany, 2008.

[87] L. Sweeney, "K-anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[88] J. Sen, "Security and privacy challenges in cognitive wireless sensor networks," in *Cognitive Radio Technology Applications for Wireless and Mobile Ad Hoc Networks*, N. Meghanathan and Y. B. Reddy, Eds., pp. 194–232, IGI-Global, Hershey, Pa, USA, 2013.

[89] Y. Ouyang, Z. Le, Y. Xu et al., "Providing anonymity in wireless sensor networks," in *Proceedings of the IEEE International Conference on Pervasive Services (ICPS '07)*, pp. 145–148, July 2007.

[90] S. Chakraborty, K. R. Raghavan, M. P. Johnson, and M. B. Srivastava, "A framework for context-aware privacy of sensor data on mobile systems," in *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, pp. 11:1–11:6, ACM, New York, NY, USA, 2013.

[91] M. Gruteser and D. Grunwald, "Anonymous usage of locationbased services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, pp. 31–42, ACM, New York, NY, USA, 2003.

[92] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *Proceedings of the ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN '04)*, pp. 88–93, October 2004.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 352421, 12 pages
http://dx.doi.org/10.1155/2014/352421

Research Article

Collection Tree Extension of Reactive Routing Protocol for Low-Power and Lossy Networks

Jiazi Yi and Thomas Clausen

Laboratoire d'Informatique-(LIX), Ecole Polytechnique, Palaiseau, 91128 Route de Saclay, France

Correspondence should be addressed to Jiazi Yi; yi.jiazi@gmail.com

Received 31 October 2013; Revised 19 February 2014; Accepted 20 February 2014; Published 25 March 2014

Academic Editor: Christos Verikoukis

This paper proposes an extension to reactive routing protocol, for efficient construction of a collection tree for data acquisition in sensor networks. The Lightweight On-Demand Ad hoc Distance Vector Routing Protocol-Next Generation (LOADng) is a reactive distance vector protocol which is intended for use in mobile ad hoc networks and low-power and lossy networks to build paths between source-destination pairs. In 2013, ITU-T has ratified the recommendation G.9903 Amendment 1, which includes LOADng in a specific normative annex for routing protocol in smart grids. The extension uses the mechanisms from LOADng, imposes minimal overhead and complexity, and enables a deployment to efficiently support "sensor-to-root" traffic, avoiding complications of unidirectional links in the collection tree. The protocol complexity, security, and interoperability are examined in detail. The simulation results show that the extension can effectively improve the efficiency of data acquisition in the network.

1. Introduction

"*The Internet of Things*" (IoT) assumes objects in our environment be part of the Internet, communicating with users and with each other and that these objects have communication as a commodity. Communication in "The Internet of Things" is a challenge, subject to resource constraints, fragile and low-capacity links, and dynamic and arbitrary topologies. Routing is among the challenges, which requires efficient protocols, able to converge rapidly even in very large networks, while exchanging limited control traffic and requiring limited memory and processing power.

One of the important applications of IoT is for data acquisition in sensor networks: a set of spatially distributed sensors that are used to monitor physical or environmental conditions, and transmit their data to a data concentrator (periodically, or triggered by some events). These data are transmitted by way of a multihop network and where the intermediary hops (routers) in that network are the sensor devices themselves. The collection of all paths from each sensor to the data concentrator forms a *collection tree*. Traffic in such a collection tree is commonly described as being "sensor-to-root" traffic or "multipoint-to-point" traffic, indicating that all traffic flows from the sensors to the data concentrator.

This paper describes a protocol for constructing such a collection tree in multihop sensor networks, where the protocol ensures that the resulting collection tree contains bidirectional paths between each sensor and the data concentrator. The protocol is defined as an extension to a reactive protocol: the LOADng routing protocol [1], which provides point-to-point routes between any two devices in a sensor network. Deploying both in unison permits efficient construction of both point-to-point routes and collection trees, by way of the same, simple protocol mechanisms.

1.1. Background and History. Since the late 90s, the Internet Engineering Task Force (IETF) (http://www.ietf.org) has embarked upon a path of developing routing protocols for networks with increasingly more fragile and low-capacity links, with less predetermined connectivity properties and with increasingly constrained router resources. In '97, the MANET (Mobile Ad hoc Networks) working group was charted, then subsequently in 2006 and 2008, 6LoWPAN (IPv6 over Low Power WPAN) and ROLL (Routing over Low Power and Lossy Networks) working groups were charted.

(1) MANET Protocol Developments. The MANET working group has developed two protocol families: reactive

protocols, including AODV (Ad hoc On-Demand Distance Vector Routing [2]), and proactive protocols, including OLSR (Optimized Link State Routing [3]). A distance vector protocol, AODV, operates in an *on-demand* fashion, acquiring and maintaining routes only while needed for carrying data, by way of a *Route Request-Route Reply* exchange. A link state protocol, OLSR, uses a periodic control messages exchanges, each router proactively maintaining a routing table with entries for all destinations in the network, which provides low delays but constant control overhead. A sizeable body of work exists, including studying the performance of these protocols in different scenarios and justifying their complementarity [4]. For the purpose of this paper, it suffices to observe that OLSR provides low delays and predictable, constant control overhead—at expense of requiring memory in each router for maintaining complete network topology. AODV limits the memory required for routing state to that for actively used routes—at the expense of delays for the *Route Request-Route Reply* exchange to take place and control overhead dependent on data flows.

After acquiring operational experiences, the MANET working group commenced by developing successors to OLSR and AODV, denoted by OLSRv2 and DYMO (Dynamic MANET On-Demand Routing). Whereas a relatively large and active community pushed OLSRv2 towards standardisation [5, 6], the momentum behind DYMO withered in the MANET working group http://tools.ietf.org/wg/manet/minutes?item=minutes81.html).

(2) 6LowPAN, ROLL, and Related Protocol Developments. The 6LowPAN (IPv6 over low power WPAN) working group was chartered for adapting IPv6 for operation over IEEE 802.15.4, accommodating characteristics of that MAC layer, and with a careful eye on resource constrained devices (memory, CPU, energy, and so forth). Part of the original charter for this working group was to develop protocols for routing in multihop topologies under such constrained conditions and over this particular MAC. Two initial philosophies to such routing were explored: *mesh-under* and *route-over*. The former, mesh-under, would, as part of an adaptation layer between 802.15.4 and IP, provide layer 2.5 multihop routing, that is, using link layer address for routing and presenting an underlying mesh-routed multihop topology as a single IP link. The latter, route-over, would expose the underlying multihop topology to the IP layer, whereupon IP routing would build multihop connectivity.

Several proposals for routing were presented in 6LowPAN, for each of these philosophies, including LOAD (6LoWPAN Ad hoc On-Demand Distance Vector Routing [7]). LOAD was a derivative of AODV but was adapted for link layer addresses and mesh-under routing, with some simplifications over AODV (e.g., removal of intermediate router replies and sequence numbers). However, 6LowPAN was addressing other issues regarding adapting IPv6 for IEEE 802.15.4, such as IP packet header compression, and solving the routing issues was suspended, delegated to a working group ROLL, and created in 2008 for this purpose. ROLL produced a routing protocol denoted by "routing protocol for low-power lossy networks" (RPL) [8] in 2011 based on the idea of collection tree protocol [9].

(3) Finally, towards LOADng. RPL as a collection tree protocol has several well-known issues with respect to supporting different kinds of traffic patterns, unidirection link handling, and algorithmic and code complexity [10]. On the other hand, while LOAD [7] development was suspended by the 6LoWPAN working group, AODV derivatives live on: IEEE 802.11s [11] is based on AODV, and the ITU-T G3-PLC standard [12], published in 2011, specifies the use of [7] at the MAC layer, for providing mesh-under routing for utility (electricity) metering networks. Justifications for using an AODV derivative in preference to RPL include that the former better supports bidirectional data flows such as a request/reply of a meter reading, as well as algorithmic and code complexity reasons [10].

The emergence of LLNs thus triggered a renewed interest in AODV-derived protocols for specific scenarios, resulting in work within the IETF [1, 13] for the purpose of standardisation of a successor to LOAD—denoted by LOADng (the Lightweight On-Demand Ad hoc Distance Vector Routing Protocol-Next Generation). LOADng incorporates the experiences from deploying LOAD—including, but not only, LLNs—and has been accepted as part of an update to the G3-PLC (Power Line Communication) ITU-T (International Telecommunication Union—Telecommunication Standardization Sector) standard for communication in the "smart grid" [14].

1.2. Statement of Purpose. There have been a lot of protocols proposed for data acquisition in sensor networks. In [15], the authors proposed collection tree protocol that uses ETX (expected transmission count) as the routing metric to construct one-way collection tree. A CDS-based network backbone for data collection is introduced in [16], to balance energy consumption and prolong the router lifetime in the backbone. A Pareto based multioptimization approach POCTP (Pareto Optimal Collection Tree Protocol) is discussed in [17] to ensure QoS such as transmission throughput, delay, and loss of packets. In [18], an average transmission time (ATT) metric is applied to routing protocol, under which real-time events are transferred along the routes with the shortest transmission time expectation. Multichannel is also used in [19] to reduce interference. Those protocols, some of them only support one-way traffic from sensor routers to one concentrator like [15], or hard to be extended for general sensor-to-sensor communications [16, 17]. Some of the protocols like [19] require specific support from lower layers, which are hard to be applied to normal sensor equipment.

The LOADng core specification aims at finding a route between any originator-destination pairs. This kind of point-to-point traffic pattern matches the basic traffic model of the Internet. However, in the world of smart grid, another important traffic pattern, called sensor-to-root, or multipoint-to-point exists. In such kind of scenarios, there are one or more concentrators that play as "root," and all the other routers

communicate with the root. If routes from all the other routers to the root are required, it is more efficient to build a "collection tree," which is a directed graph that all edges are oriented toward and terminate at one root router.

This paper proposes an extension to a reactive routing protocol LOADng, denoted by LOADng Collection Tree Protocol (LOADng-CTP), for building a "collection tree" in environments, constrained in terms of computational power, memory, and energy. An example of the design target for LOADng-CTP is the ESB (Embedded Sensor Board [20]), with a TI MSP430 low-power microcontroller, an 1 MHz CPU, 2 kB RAM, and 60 kB flash ROM. The link layers typically used in LLNs impose strict limitations on packet sizes: in IEEE 802.15.4, the maximum physical layer packet size is 127 bytes and the resulting maximum frame size at the mac-layer is 102 bytes. If link-layer security is used, this may consume up to a further 21 bytes, which leaves just 81 bytes for upper layer protocols.

The LOADng-CTP presented in this paper is thus designed to meet the following requirements:

(i) effectively building a route from all sensors to the root and the route from the root to the sensors if required;

(ii) unidirectional links being avoided in these routes;

(iii) low overhead, easy collection tree maintenance;

(iv) easy extension to LOADng, such that routers using only LOADng (without collection tree extension) can join the collection tree.

Although the specification in this paper is designated for LOADng, its basic mechanism can be applied to general reactive protocol, like AODV also.

The remainder of this paper is organized as follows. In Section 2, the LOADng-CTP specification is introduced, including related message format and main operations. The protocol is further analysed in Section 3, from the aspect of routing complexity, security, and interoperability. The simulation study is performed in Section 4, in which LOADng, LOADng-CTP, and RPL are compared. Section 5 concludes this paper.

2. LOADng-CTP Protocol Specification

LOADng Collection Tree Protocol (LOADng-CTP) is based on the operation and packet format of LOADng. Therefore, the current LOADng implementation can be easily extended to the collection tree protocol. In the following, the basic operation of LOADng is introduced briefly, followed by the single message and protocol processing required for collection tree building and maintenance.

2.1. LOADng Basic Operation. LOADng contains two main operations: *Route Discovery* and *Route Maintenance*.

(1) Route Discovery. During *Route Discovery*, RREQ (Route Request) messages are flooded through the network. In LOADng [1], only the destination of the RREQ will reply by generating and unicasting an RREP (Route Reply) to the originator of the RREQ. All RREQ and RREP messages, generated by a LOADng router, carry a monotonically increasing sequence number, permitting both duplicate detection, and detecting which of two messages contain the most "fresh" information.

(2) Route Maintenance. Route Maintenance is performed when an actively used route fails. Route failure is detected by way of a data packet not being deliverable to the next hop towards the intended destination. In LOADng, the RERR is unicasted to the source of data packet. On receiving the RERR at the source of data packet, a new Route Discovery can be performed, in order to discover a new route to the intended destination.

Compared to AODV, LOADng has the following characteristics.

(i) Modular design: the core specification defines the simple and lightweight core functions of the protocol. LOADng is extensible, by way of a flexible packet format permitting addition of arbitrary attributes and information via new message types and/or TLV (type-length-value) blocks. The LOADng protocol core is detailed in this section, with subsequent sections illustrating the use of the flexible architecture of LOADng for developing (interoperable and backwards compatible) protocol extensions.

(ii) Optimised flooding: It can reduce the overhead incurred by RREQ forwarding. Jitter is employed, to reduce the probability of losses due to collisions on lower layers [21].

(iii) Flexible addressing: address lengths from 1 to 16 octets are supported (i.e., IPv6, IPv4, 6LowPAN short addresses, Layer-2 MAC addresses, and so forth are all supported by LOADng). The only requirement is that, within a given routing domain, all addresses are of the same address length.

(iv) Metrics: different metrics are supported, to make use of link information from different layers.

(v) Destination replies: intermediate LOADng routers are explicitly prohibited from responding to RREQs, even if they may have active routes to the sought destination. All messages (RREQ or RREPs) generated by a given LOADng router share a single unique, monotonically increasing sequence number. This also eliminates Gratuitous RREPs while ensuring loop freedom. The rationale for this simplification reduced complexity of protocol operation and reduced message sizes—found to be without significant influence in the performance [22]. Allowing only the destination to reply to an RREQ also simplifies the task of securing the protocol, because the destination can thus sign the RREP message, and the originator could verify that it is the "real" destination that replies.

(vi) Reduced state: a LOADng router is not required to maintain a precursor list; thus when forwarding a data packet to the recorded next hop on the path to the destination fails, an RERR is sent only to

the originator of that data packet. The rationale for this simplification is an assumption that few overlapping routes are in use concurrently, and delay is not a critical issue in a given network.

2.2. Message for LOADng-CTP. LOADng-CTP introduces two flags to RREQ messages, carried by a so-called RREQ flag.

(i) RREQ COLLECTION_TREE_TRIGGER: when set, a receiving router will be triggered to discover with which of its neighbours it has bidirectional links.

(ii) RREQ COLLECTION_TREE_BUILD: when set, a receiving router will build the route to the root.

In addition, a HELLO message [5] is used, which includes all the 1-hop neighbours of the router generating the HELLO message. The HELLO messages are broadcast and never forwarded. Bidirectional neighbours are identified by the exchange of HELLO messages.

2.3. Router Parameters for LOADng-CTP. LOADng-CTP uses the following parameters for protocol functioning.

(i) NET_TRAVERSAL_TIME: it is the maximum time that a packet is expected to take when traversing from one end of the network to the other.

(ii) RREQ_MAX_JITTER: it is the maximum jitter for RREQ message transmission. Jitter is a randomly modifying timing mechanism to control traffic transmission in wireless networks to reduce the probability of transmission collisions [21].

(iii) HELLO_MIN_JITTER: it is the minimum jitter for HELLO message transmission. HELLO_MIN_ JITTER must be greater than 2 × RREQ_MAX_JITTER.

(iv) HELLO_MAX_JITTER: it is the maximum jitter for HELLO message transmission.

(v) RREP_REQUIRED: it is the flag to define if an RREP message is required on receiving RREQ_BUILD message, to build routes from the root to sensors.

2.4. LOADng-CTP Procedures. The collection tree is, then, built by way of the following procedure—initiated by the router wishing to be the root of the collection tree.

(1) Collection Tree Triggering (by the Root). The root generates an RREQ with *COLLECTION_TREE_TRIGGER* set (henceforth, denoted by RREQ_TRIGGER). Both the originator and destination of the RREQ_TRIGGER are set to the address of the root.

When an RREQ_TRIGGER is generated, an RREQ with COLLECTION_TREE_BUILD flag set (henceforth, denoted by RREQ_BUILD) is scheduled to be generated in 2 × NET_TRAVERSAL_TIME.

(2) Bidirectional Neighbour Discovery. On receiving a RREQ_TRIGGER, a router does the following.

(i) It records the address of the sending router (i.e., the neighbour, from which it received the RREQ_TRIGGER) in its *neighbour set*, with the status *HEARD*.

(ii) If no earlier copy of that same RREQ_TRIGGER has been previously received,

(a) the RREQ_TRIGGER is retransmitted, subject to a jitter of RREQ_MAX_JITTER, to reduce the chance of collisions (except the root router);

(b) it schedules generation of a HELLO message, subject to a jitter of between HELLO_MIN_ JITTER and HELLO_MAX_JITTER. When the scheduled HELLO message is generated, it lists the addresses of all the 1-hop neighbours, from which it has received a RREQ_TRIGGER.

On receiving a HELLO message, a router does the following.

(i) If it finds its own address listed in the HELLO message, it records the address of the sending router (i.e., the neighbour, from which it received the HELLO) in its *neighbour set*, with the status *SYM* (bidirectional).

(ii) The HELLO message is never forwarded but discarded silently.

Thus, each router will learn with which among its neighbour routers it has a bidirectional (SYM) or unidirectional (HEARD) link.

(3) Collection Tree Building. For 2 × NET_TRAVERSAL_ TIME after the RREQ_TRIGGER, the root generates a RREQ_BUILD.

On receiving a RREQ_BUILD, a router does the following.

(i) It verifies if the RREQ_BUILD was received from a neighbour with which it has a bidirectional (SYM) link. If not, the RREQ_BUILD is silently discarded.

(ii) Otherwise, if no earlier copy of that same RREQ_BUILD has been previously received, or the RREQ_BUILD indicates a short path to the root,

(a) a new routing entry is inserted into the routing table, with

(1) *next_hop* = previous hop of the RREQ_BUILD,

(2) *destination* = root;

(b) the RREQ_BUILD is retransmitted, again subject to a jitter of RREQ_JITTER.

Thus, each router will record a route to the root, and this route will contain only bidirectional links. The collection tree is built, enabling upward traffic. Figure 1 illustrates the RREQ_BUILD processing.

(4) Root-to-Sensor Path Building. By exchanging RREQ_ TRIGGER and RREQ_BUILD messages, all the sensors in the network obtained a path using only bidirectional links to

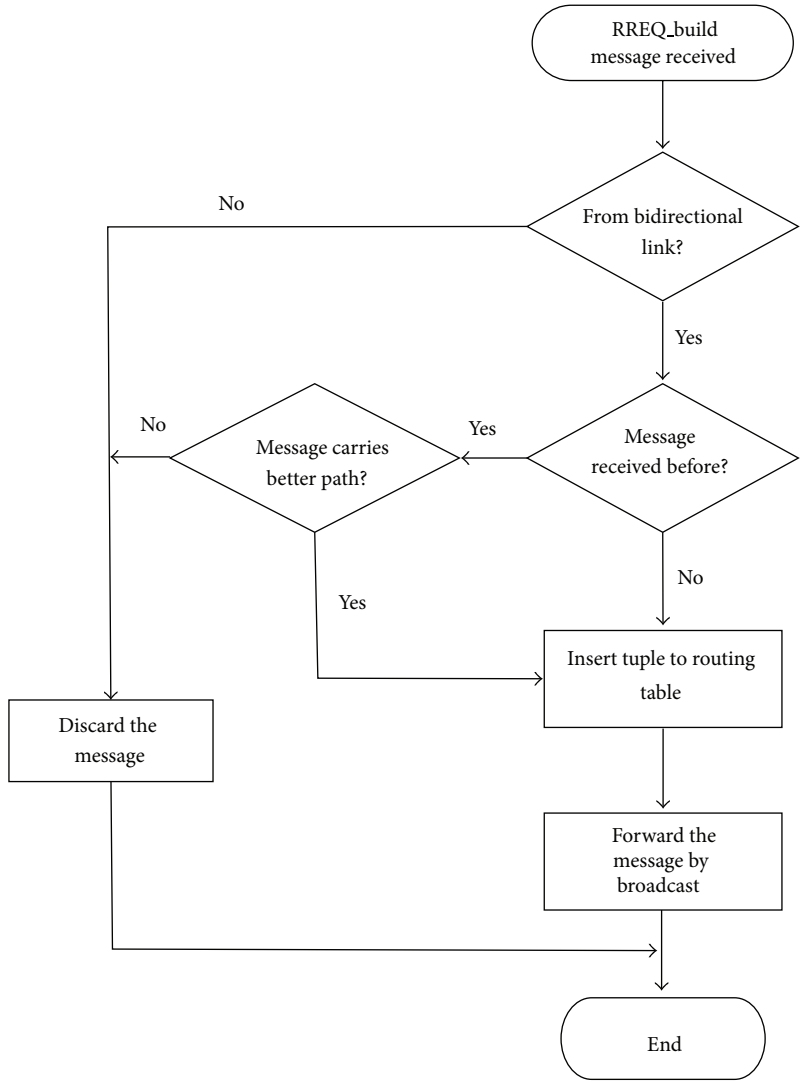


Figure 1: LOADng-CTP RREQ_BUILD message processing.

the root. This is sufficient for applications like environment monitoring, automatic meter reading, and so forth. However, in some applications, such as firmware update or remote control, the root needs to send messages to sensors in the network. The paths from root to sensors are thus desired.

The sensors that require root-to-sensor traffic must have their RREP_REQUIRED flag set to be true. On receiving the RREQ_BUILD message, all the sensor routers with RREP_REQUIRED flag set must initiate an RREP message with content of

(i) *RREP_originator* = address of the sensor router;

(ii) *RREP_destination* = address of the root.

The RREP is thus unicast to the root, subject to jitter RREP_JITTER. On receiving the RREP message, a routing tuple is created in the routing table with

(i) *next_hop* = previous hop of the RREP;

(ii) *destination* = address of the RREP originator (*RREP_originator*).

Figure 2 depicts an example of root-sensor message exchange sequences by illustrating the four steps of LOADng-CTP protocol (collection tree triggering, bidirectional neighbour discovery, collection tree build, and root-to-sensor path building). In the example, the *root* router builds a collections tree connecting sensor routers *A* and *B*, with the topology shown in Figure 2(a). The message exchange is shown in Figure 2(b). The pseudosequence number in the brackets is used just for distinguishing different messages in this figure. In a real protocol implementation, sequence numbers are generated independently at each router.

2.5. Collection Tree Maintenance. Based on the operation introduced in Section 2.4, a collection tree is built to enable data traffic transmission between the root router and all the other sensors. However, route failure could still happen,

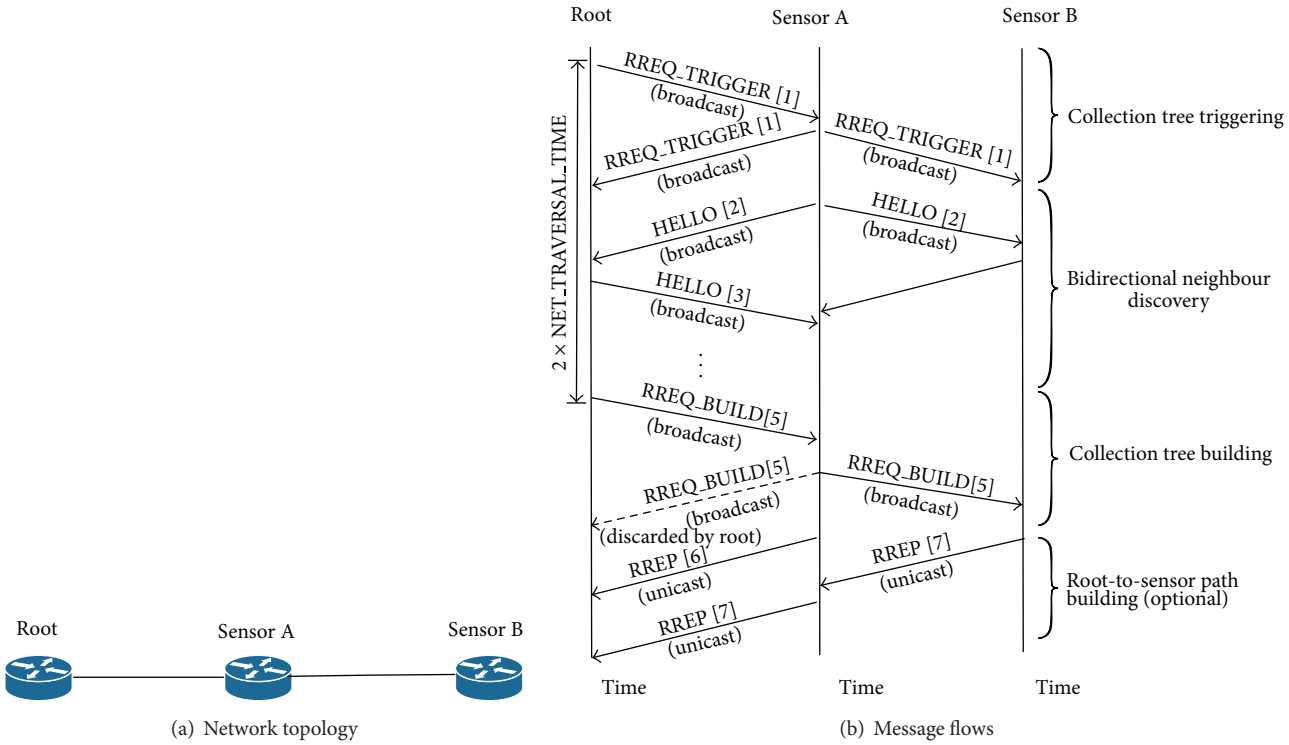
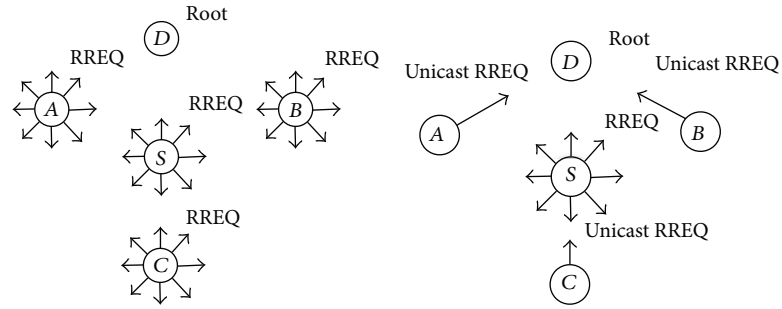


Figure 2: Message exchange of LOADng-CTP between root and sensors.



(a) Route discovery without smart route request

(b) Route discovery with smart route request

Figure 3: An example of route maintenance. Router *D* is the root. The link between *S*–*D* is detected broken. Sensor routers *A, B, and C* still have routing tuple to *D*.

due to the "lossy" nature of sensor networks or topology changes, such as

(i) loss of control message during the collection tree building process;

(ii) routing entries expire because of not updated timely;

(iii) participation of new sensors;

(iv) Sensors quit the network because of movement or battery drain.

LOADng-CTP supports per-path maintenance when a path failure is detected, without rebuilding the whole collection tree. A new route discovery is initiated according to usual procedures of route discovery if

(i) the data packet to be forwarded cannot find a routing tuple to the desired destination in the routing table, or

(ii) the link to the "next hop" indicated by the routing table is detected broken.

To avoid RREQ being broadcast through the whole network and take benefits from that "most of other neighbour routers might have an available route to the root," a *Smart Route Request* scheme can be employed: if an intermediate router, receiving the RREQ, does not have an available route to the destination, the RREQ is forwarded as normal. If the intermediate router has a route to the root, that intermediate router will unicast the RREQ to the destination according to the routing table.

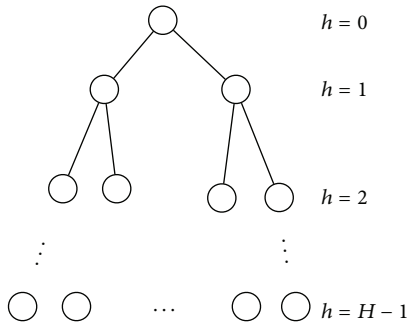


Figure 4: An example of balanced tree. Every parent has 2 children ($C = 2$).

Figure 3 gives an example of path maintenance in collection tree. Router D is the root, and the link between S–D is detected broken. Routers A and B have already direct path to D, and C has also a routing tuple to D (by going through S). Figure 3(a) depicts the route discovery initiated by S according to LOADng basic operation. Because only the destination is allowed to reply to the RREQ message, sensor routers *A, B, and C* have to rebroadcast the RREQ message, even they have already routing tuples to D. This renders a network-wide flooding: for a network with n routers, n RREQ message retransmissions are required.

With smart route request, as shown in Figure 3(b), routers *A, B, and C* will unicast the RREQ to root D according to their routing tables, and D can choose the best path to send RREP message. By doing so, the RREQ dissemination is limited locally (4 retransmission in this example), and the routing overhead can be greatly reduced.

When a link on an active route to a destination is detected as broken (by way of inability to forward a data packet towards that destination), an RERR (route error) message is unicast to the source of the undeliverable data packet. Both this intermediate router and the source router need to initiate a new route discovery procedure.

3. LOADng-CTP Protocol Analyses

This section analyzes the main features of the LOADng-CTP, including protocol complexity, security considerations, and its interoperability with LOADng protocol.

3.1. Protocol Complexity. Unlike link-state routing protocols such as OSPF [23] or OLSR [6], which require keeping a network topology locally and run the Dijkstra algorithm, LOADng and LOADng-CTP concern only the basic additive operation when calculating link metrics. Therefore, the computational complexity is negligible. A very important concern of routing protocol for sensor networks is its routing overhead: the message required to maintain the routing table.

For simplicity, a balanced tree model is considered: there is a single root in the tree, with total height of H. The height of root is 0, and the leaf nodes are with height $H-1$. Every node in the tree (except the leaf nodes) has C children ($C > 1$). Figure 4 gives an example of balanced tree with $C = 2$.

The number of nodes at height h ($0 \le h \le H-1$) is $n_h = C^h$. The total number of nodes in the tree is

$$N = 1 + C + C^2 + \cdots + C^{(H-1)} = \frac{1 - C^H}{1 - C} \quad (C > 1). \tag{1}$$

In LOADng-CTP, the message required for collection tree building is the sum of RREQ_TRIGGER, HELLO, and RREQ_BUILD:

$$\text{RREQ} = 3N. \tag{2}$$

If root-to-sensor paths are required, every sensor also has to unicast an RREP message to the root.

The number of RREP messages forwarded by all the routers at height h is

$$\text{RREP}_h = C^h \sum_{i=0}^{H-h-1} C^i = C^h \frac{1 - C^{H-h}}{1 - C}. \tag{3}$$

The total number of RREP can thus be given by

$$\text{RREP}_{\text{All}} = \sum_{h=1}^{H-1} \text{RREP}_h = \sum_{h=1}^{H-1} \frac{C^h}{1 - C} - \sum_{h=1}^{H-1} \frac{C^H}{1 - C}. \tag{4}$$

Considering (1), the total number of RREP forwarding is

$$\begin{aligned} \text{RREP}_{\text{All}} &= \frac{1}{1 - C} \frac{C - C^H}{1 - C} - \frac{(H-1)\, C^H}{1 - C} \\ &= NH - N + \frac{N - H}{1 - C}. \end{aligned} \tag{5}$$

Considering $H = \lfloor \log_C N \rfloor$, the total number of RREP messages thus scales with $O(N \log N)$.

For the basic LOADng protocol, by which only point-to-point route build is supported, the number of RREQ messages forwarding required to build path from all the sensors to the root is

$$\text{RREQ} = N^2. \tag{6}$$

The RREP message is always needed in LOADng basic operation, which is the same as (5).

Based on (2), (5), and (6), it can be concluded that LOADng-CTP reduced routing overhead from $O(N^2)$ to $O(N)$ compared to basic LOADng mechanism if only sensor-to-root paths are needed, or $O(N \log N)$, if root-to-sensor paths are also required.

3.2. Security Considerations

(1) Protocol Vulnerability. The collection tree building process relies on strictly ordered message sequences: RREQ_TRIGGER message for triggering the building process, then HELLO message for bidirectional neighbour check, and RREQ_BUILD message for collection tree build in the end. The message emission is controlled by router parameters like NET_TRAVERSAL_TIME, RREQ_JITTER, and HELLO_JITTER.

The receiving order can be expected if those parameters are set correctly; however, in real implementations, there might exist misconfigured routers, or even compromised routers that emit messages out of order. For example, if a router sends a HELLO message before it receives all the RREQ_TRIGGER messages from its neighbours, or an RREQ_BUILD message is received before the HELLO message exchange finished, the router cannot identify its bidirectional neighbours correctly—thus it is not able to join the collection tree as expected.

In addition to message misordering, LOADng-CTP is also prone to attacks like block-hole or spoofing attacks [24, 25]. Malicious control traffic can have severe impact on the network stability.

(2) Security Framework. One of the main objectives when specifying LOADng was to provide a modular architecture with a core module that is easily extensible. The rationale for this decision was that rarely "one-size-fits-all" in the area of constrained networks. This is particularly true for security extensions: some networks may not require any level of Layer 3 security, for example, because physical access is limited or lower layer protection is sufficient. Other networks require integrity protection with a lightweight cipher suite due to limited processing power and memory of routers. In some cases, security requirements are tighter and confidentiality as well as strong cryptographic ciphers is required.

The IETF has standardized a security framework for protocols using the message and packet format defined in [26], which is used by LOADng-CTP. (Note that this framework is currently being revised in a succeeding document that will obsolete RFC6622 once approved: http://tools.ietf.org/html/draft-ietf-manet-rfc6622-bis). Reference [27] specifies a syntactical representation of security-related information in TLVs for use with [26] addresses, messages, and packets. That specification does not represent a stand-alone protocol but is intended for use by MANET routing protocols, or security extensions thereof, such as LOADng-CTP.

Figure 5 depicts the architecture of a module for LOADng-CTP that provides integrity and nonrepudiation for LOADng, using the framework specified in [27].

Incoming RFC5444 packets are first parsed by the RFC5444 parser that demultiplexes messages and sends them to the protocol "owning" the message type. As each RFC5444 packet may contain multiple messages that are used by different protocols on a router, the message type is used to demultiplex and send the message to the appropriate protocol instance. A message intended for LOADng-CTP will then be forwarded to the security extension module that verifies the signature contained in a signature TLV inside the message. As the TLV contains additional information, such as the hash function (e.g., SHA-256, Secure Hash Algorithm) and the cryptographic function (e.g., AES, Advanced Encryption Standard), the module can choose the correct key and verify the integrity protection. If the message signature is correct, the message is handed over to the LOADng-CTP module; otherwise it is rejected. Similarly, outgoing messages from LOADng-CTP are handed over to the security module, which

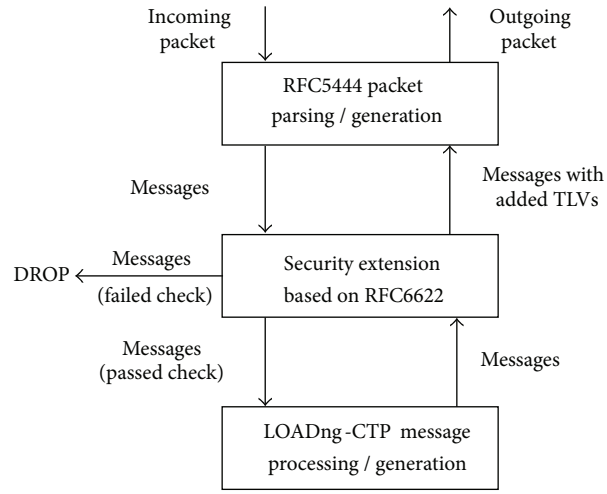


Figure 5: Relationship with RFC5444, RFC6622, and LOADng-CTP.

in turn adds the TLV containing the digital signature of the message. Then the message is handed over to the RFC5444 module that multiplexes it into a packet.

During the message signature generation as well as verification process, [27] takes special consideration for mutable fields, such as hop count and hop limit. In addition to hop count and limit, the route metric contained in a metric TLV is also updated along the path of a message and can therefore not be protected by a digital signature. LOADng-CTP lists these mutable fields explicitly. While this is a security problem that needs to be addressed in addition to a pure message signature (and is not discussed in this paper), based on the message format of LOADng-CTP messages, at least the calculation of signature is easy. This is because the message size does not change as no field is added or removed during the forwarding process of a message through the network (and therefore no other fields, such as message size or TLV block size, need to be recalculated). The metric can simply be replaced by a sequence of zeros before calculating the signature and is then restored afterwards.

In addition to message integrity, packets may also be digitally signed. As packets are used hop-by-hop, that is, are never forwarded, this is useful to authenticate the previous hop along the path of a message. Otherwise, a router not having any credentials may, for example, simply forward a correctly signed RREP message from one adjacent router to another and increase the hop count. As the hop count is excluded from the signature calculation, the message integrity would still be valid. Packet signatures mitigate this problem at the expense of increased overhead on the channel. Note also that it is difficult to detect simple forwarding of a frame without modifying the content, also known as "wormhole attack."

3.3. Interoperability Considerations. As sensor networks and low-power and lossy networks are generally decentralized system, devices would possibly work in a heterogeneous environment: there might be old devices with basic functions,

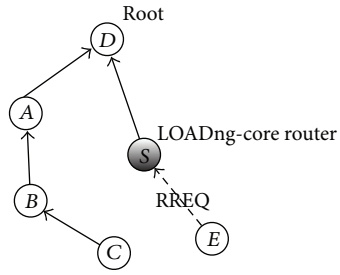


Figure 6: An example of interoperability between LOADng-CTP (white nodes) and LOADng-core routers (grey nodes).

and newly jointed devices with extensions in the same routing domain. This requires interoperability between routers using LOADng-CTP and LOADng routers without collection tree extension (denoted by LOADng-core router).

A LOADng-core router will forward RREQ_TRIGGER and RREQ_BUILD message as normal RREQ messages, so it will not affect the collection tree building process of other routers in the network. But because LOADng-core routers cannot generate HELLO messages themselves and are not able to be verified as bidirectional neighbour, therefore, LOADng routers will not join the collection tree during the collection tree building process described in this section. However, these routers can participate in the collection tree by initiating a new RREQ message to the root and thus join the collection tree as "leaf nodes" (i.e., nodes without children), as shown in Figure 6.

During the collection tree building process, LOADng-core routers will not be able to function as parents of other routers. As depicted in Figure 6, router C will choose B as parent, even if S probably provides a shorter path to the root. If the LOADng-core router is on the only path to the root,for example, router E has to go through S to reach the root, a new RREQ will be initiated to the root.

The existence of LOADng-core routers will possibly increase the routing overhead in the network by initiating more route discoveries. But with the smart RREQ introduced in Section 2.5, the RREQ dissemination can be kept locally, thus without introducing much influence in the networks.

4. Simulation and Performance Analyses

4.1. Simulation Settings. In order to understand the performance impact of the collection tree extension to LOADng, this section presents a set of ns2 simulations, comparing LOADng, LOADng-CTP, and RPL, with the parameters of the trickle timer in RPL being set according to [8]. Simulations were made with varying numbers of routers from 63 to 500 and placed statically randomly in a square field. The networks have consistent density of nodes; that is, the simulation field grows as the number of routers increases: 1100 m × 1100 m for 63 nodes, 1580 m × 1580 m for 125 nodes, 2230 m × 2230 m for 250 nodes, and 3160 m × 3160 m for 500 nodes. This simulates smart grid in suburban areas. As the size of the network grows, the scalability of the protocol can be tested.

The network is subject to sensor-to-root traffic, like periodic meter reading: all routers generate traffic, for which the destination always is a single, fixed router in the network. Each data source transmits a 512-byte data packet every 5 seconds, in bursts lasting for 80 seconds each, for a total simulation time of 100 s.

For the purpose of this study, router mobility was not considered. Simulations were conducted using the TwoRay-Ground propagation model and the IEEE 802.11 MAC. Although there are various low-layer technologies more commonly (and, perhaps, more viably) used for LLNs (power line communication, 802.15.4, low-power wifi, Bluetooth low energy, etc.), 802.11 provides basic distributed mechanisms for channel access, such as DCF (distributed coordination function), CSMA/CA (carrier sense multiple access with collision avoidance). Therefore, general behaviour of a protocol can be inferred from simulations using 802.11.

In the simulations, three types of routing protocols are compared.

(i) LOADng core specification [1], referred to as LOADng in the following section. The routes are built reactively when there are data packets need to be sent.

(ii) LOADng with collection tree extension, referred to as LOADng-CTP. The collection tree is triggered and built before the sending of data packets.

(iii) RPL with trickle timer, referred to as RPL. The parameters of trickle timer are set according to [8].

4.2. Simulation Results. Figure 7 depicts the delivery ratio of three protocols. Both LOADng-CTP and RPL obtain delivery ratios close to 100%, regardless of number of nodes. LOADng, initiating route discovery for every router (network-wide broadcast), incurs a high number of collisions on the MAC layer (shown in Figure 8), and thus a lower data delivery ratio, especially in larger scenarios.

Figure 9 illustrates the average end-to-end delay. LOADng has longer delay mainly because the route discovery is performed reactively; that is, the data packets have to wait the finish of route discovery before being sent out. LOADng-CTP and RPL have routes a priori available, thus exhibiting identical delays.

For the sensor networks, the routing overhead is also a crucial consideration. Figures 10 and 11 show the number of overhead packets per router and average overhead of network (bytes/second), respectively, which the networks are needed to converge to a stable state; that is, every router has a route to the root.

The overhead packets of LOADng-CTP and RPL grow linearly with RPL sending twice as many packets as LOADng-CTP and RPL sending 10 times more bytes/s as compared to LOADng-CTP, due to the RPL control packets (mainly, the DIOs) being bigger [10]: a DIO packet takes up to 40 octets in these scenarios, whereas a LOADng-CTP RREQ and RREP packet typically are 10 octets (base header of 24 octets, plus other options and addresses). The overhead of LOADng grows exponentially as the number of nodes increases, up to

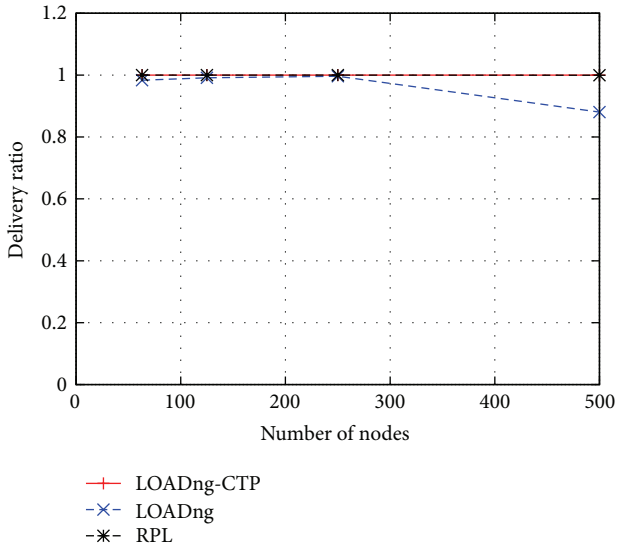


Figure 7: Packet delivery ratio.

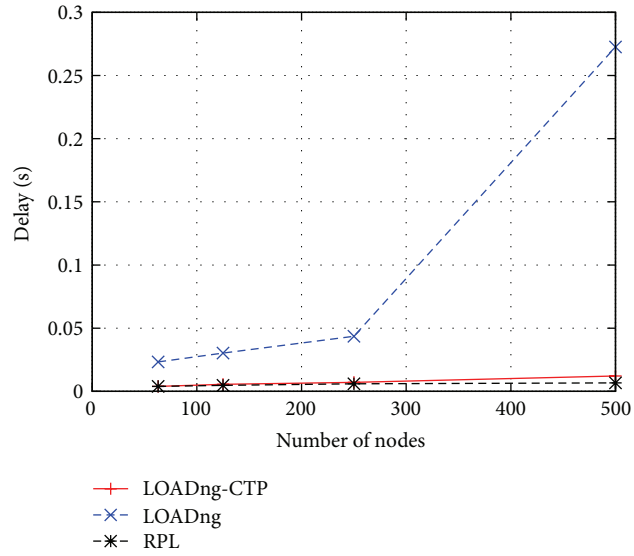


Figure 9: Average end-to-end delay.

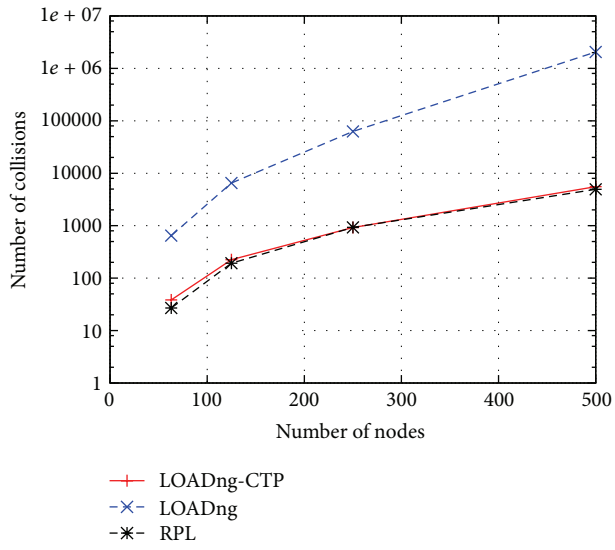


Figure 8: Number of MAC layer collisions.

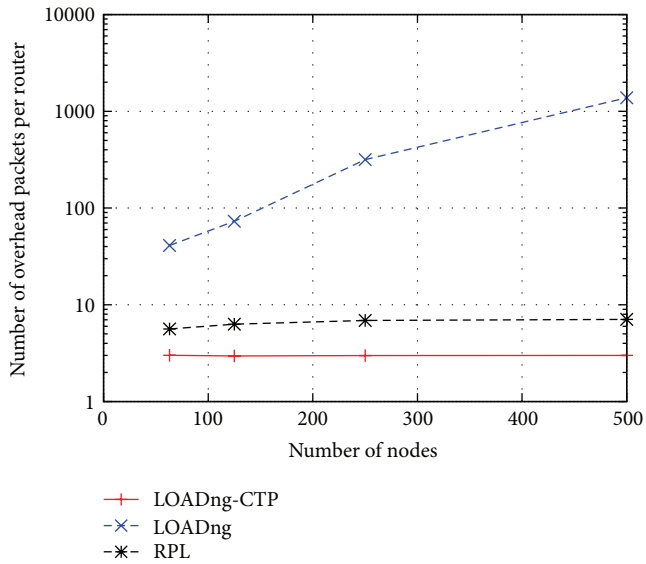


Figure 10: Number of overhead packets transmitted by each router.

700,000 packets for scenarios of 500 nodes (not drawn in the figure). The point-to-point based basic LOADng mechanism is not optimized for sensor-to-root traffic.

5. Conclusion

This paper has presented an extension, LOADng-CTP, to the reactive LOADng routing protocol, permitting efficient and on-demand construction of collection trees for supporting sensor-to-root traffic types. LOADng-CTP permits finding paths between a root router and all the other sensor routers in the network using bidirectional links. The protocol supports per-path route maintenance without rebuilding the whole collection tree. Another key aspect of LOADng-CTP is that any router can at any time determine that it needs to act as a root for sensor-to-root traffic and spawn a collection tree construction; this, without requiring that said router is specifically provisioned for this purpose (no extra state, processing power, required).

The main features of LOADng-CTP are analysed. The routing overhead is reduced to $O(N)$ for collection tree building, compared to $O(N^2)$ of LOADng core specification (N is the number of routers in the network). An extensible security framework is proposed to protect the integrity of routing message exchange. The interoperability between collection tree extension and LOADng core specification is considered. The LOADng routers without collection tree extension can also join the collection tree by initiating a route discovery.

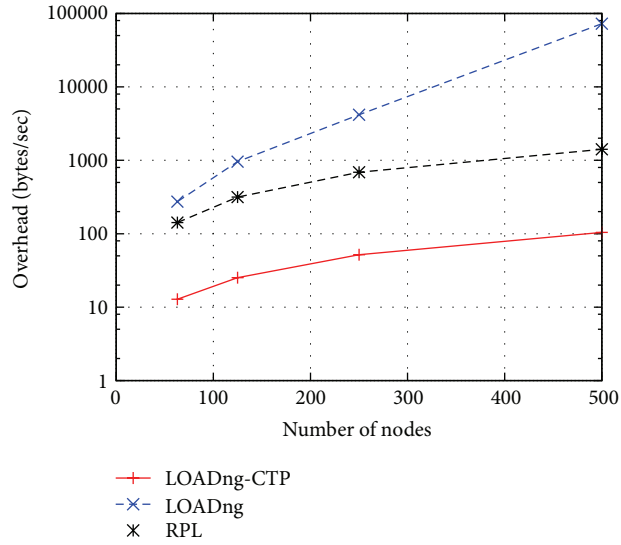


Figure 11: Overhead bytes per second in the whole network.

The performance of this extension has been studied; revealing delays and data delivery rations, comparable with RPL, are obtained while at the same time yielding considerably lower control traffic overheads. Compared to basic LOADng, the performance of the LOADng-CTP extension yields better performance: lower overhead, higher data delivery ratios, and lower delays.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to gratefully acknowledge Ulrich Herberg, Axel Colin de Verdiere, Yuichi Igarashi, Christos Verikoukis (assigned editor), and anonymous reviewers of the paper for valuable comments, reviews, and technical discussions.

References

[1] T. Clausen, A. C. de Verdiere, J. Yi et al., "The lln on-demand ad hoc distance-vector routing protocol—next generation," The Internet Engineering Task Force, October 2013, internet Draft, work in progress, draft-clausen-lln-loadng.

[2] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," Experimental RFC, 3561, July 2003.

[3] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," Experimental RFC, 3626, October 2003.

[4] L. V. T. Clausen and P. Jacquet, "Comparative study of routing protocols for mobile ad-hoc networks," in *Proceedings of the IFIP MedHocNet*, Sardinia, Italy, September 2002.

[5] T. Clausen, C. Dearlove, and J. Dean, "Mobile Ad Hoc Network (MANET) Neighborhood Discovery Protocol (NHDP)," RFC, 6130, IETF, April 2010.

[6] T. Clausen, C. Dearlove, P. Jacquet, and U. Herberg, "The Optimized Link State Routing Protocol version 2," Internet Draft, draft-ietf-manetolsrv2- 19, work in progress, March 2013.

[7] K. Kim, S. D. Park, G. Montenegro, S. Yoo, and N. Kushalnagar, "6LoWPAN Ad Hoc On-Demand Distance Vector Routing," Internet Draft, work in progress, draft-daniel-6lowpan-load-adhocrouting- 03, June 2007.

[8] T. Winter, P. Thubert, A. Brandt et al., "RPL: IPv6 Routing Protocol for Low power and Lossy Networks," IETF RFC6550, March 2011.

[9] O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis, "The Collection Tree Protocol (CTP)," TinyOS, Tech. Rep., 2009, http://sing.stanford.edu/pubs/sing-09-01.pdf.

[10] T. Clausen, A. C. de Verdiere, J. Yi, U. Herberg, and Y. Igarashi, "Experiences with RPL: IPv6 Routing Protocol for Low power and Lossy Networks," The Internet Engineering Task Force, internet Draft, work in progress, draft-clausen-lln-rpl-experiences, February 2013.

[11] G. R. Hiertz, S. Max, Z. Rui, D. Denteneer, and L. Berlemann, "Principles of IEEE 802.11s," in *Proceedings of the 16th International Conference on Computer Communications and Networks (ICCCN '07)*, pp. 1002–1007, Honolulu, Hawaii, USA, August 2007.

[12] "ITU-T G. 9956: Narrow-Band OFDM power line communication transceivers—Data link layer specification," November 2011.

[13] T. Clausen, A. Camacho, J. Yi et al., "Interoperability report for the lightweight on-demand ad hoc distance-vector routing protocol—next generation (loadng)," The Internet Engineering Task Force, internet Draft, work in progress, draft-lavenu-lln-loadng-interoperability-report, December 2012.

[14] ITU, "ITU-T G. 9903: Narrow-band orthogonal frequency division multiplexing power line communication transceivers for G3-PLC networks: Amendment 1," May 2013.

[15] J. J. Lei, T. Park, and G. I. Kwon, "A reliable data collection protocol based on erasure-resilient code in asymmetric wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 730819, 8 pages, 2013.

[16] X. Kui, Y. Sheng, H. Du, and J. Liang, "Constructing a cds-based network backbone for data collection in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 258081, 12 pages, 2013.

[17] Y.-Z. Wu, D.-P. Quan, and H.-G. Han, "Pareto Optimal Collection Tree Protocol for industrial monitoring WSNs," in *Proceedings of the IEEE GLOBECOM Workshops (GC Wkshps '11)*, pp. 508–512, Houston, Tex, USA, December 2011.

[18] Y. Tang, C. Bu, and A. Fan, "A transmission time based routing protocol for clustered collection tree wireless sensor networks," in *Proceedings of the International Conference on Computational and Information Sciences (ICCIS '10)*, pp. 21–24, Chengdu, China, December 2010.

[19] C. Buengbon, C. Tanwongvarl, and S. Chantaraskul, "Multi-channel collection tree protocol for wireless sensor networks," in *Proceedings of the 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON '13)*, pp. 1–5, 2013.

[20] "The ESB Embedded Sensor Board," 2013, http://contiki.sourceforge.net/docs/2.6/a01781.html.

[21] T. Clausen, C. Dearlove, and B. Adamson, "Jitter Considerations in MANETs," IETF Inf. RFC, 5148, February 2008.

[22] T. Clausen, J. Yi, and A. C. de Verdiere, "LOADng: Towards AODV Version 2," in VTC Fall. IEEE, pp. 1–5, 2012.

[23] J. Moy, "OSPF Version 2," RFC, 2328, IETF, April 1998.

[24] W. Wang, Y. Lu, and B. K. Bhargava, "On vulnerability and protection of ad hoc on-demand distance vector protocol," in *Proceedings of the 10th International Conference on Telecommunications (ICT '03)*, vol. 1, pp. 375–382, 2003.

[25] P. Ning and K. Sun, "How to misuse AODV: a case study of insider attacks against mobile ad-hoc routing protocols," *Ad Hoc Networks*, vol. 3, no. 6, pp. 795–819, 2005.

[26] T. Clausen, C. Dearlove, J. Dean, and C. Adjih, "Generalized Mobile Ad Hoc Network (MANET) Packet/Message Format," RFC, 5444, IETF, Feburary 2009.

[27] U. Herberg and T. Clausen, "Integrity Check Value and Timestamp TLV Definitions for Mobile Ad Hoc Networks (MANETs)," RFC, 6622, IETF, May 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 151564, 7 pages
http://dx.doi.org/10.1155/2014/151564

Research Article

Design and Experiment Analysis of a Hadoop-Based Video Transcoding System for Next-Generation Wireless Sensor Networks

Haoyu Xu,[1] Liangyou Wang,[1,2] and Huang Xie[3]

[1] *Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China*
[2] *School of Electronics and Information, Tongji University, Shanghai 201804, China*
[3] *School of Software Engineering, University of Science and Technology of China, Suzhou, Jiangsu 215123, China*

Correspondence should be addressed to Haoyu Xu; xuhaoyu@gmail.com

Received 31 October 2013; Accepted 23 December 2013; Published 24 March 2014

Academic Editor: Zuqing Zhu

The next-generation wireless sensor network (WSN) has the capability of carrying hundreds of high-definition video streams, beside the feature of massive employment of energy-efficient nodes. However, several challenges are identified with respect to the video bearing, such as the different video formats, enormous size of "raw" video, and compatibility with heterogeneous terminal devices. The video transcoding system (VTS) is widely believed to address these challenges. This paper introduces a cloud-based, more specifically, Hadoop-based, video transcoding system to fulfill the vision of bearing hundreds of HD video streams in the next generation WSN, with a discussion on optimization of several significant parameters. This paper obtains three remarkable results: (1) there is an optimal value of the number of Mappers; (2) the optimal value is closely related to the file size; (3) the transcoding time depends principally on the duration of video files rather than their sizes.

1. Introduction

The next-generation wireless sensor network (WSN), in the broad sense, has two significant features. One is the massive employment of energy-efficient nodes that extends the networks' working time from months to years. The other one is the dramatically increased throughput of the networks, thanks to the emergency of new wireless transmission technology operating on the microwave spectrum. IEEE 802.11ac is a case in point [1].

The increased link throughput, up to 1 Gbps, makes the next-generation WSN capable of carrying hundreds of high-definition video streams as cameras are deemed as one kind of optical sensors. This capability has been anticipated for a long time by many applications. In the future battlefield, thousands of mobile objects, such as helicopters, vehicles, tanks, and soldiers, will be installed with high-definition or standard-definition cameras. The next-generation WSN is suitable for transmission of videos captured by these cameras back to the headquarter. Videos are going to be stored, managed, analyzed, and even redistributed to the related people. Similar application scenarios include large scale emergent rescue during natural disasters, security level scale up for a critical event, and prompt treatment to riots or chaos and so on.

To fully and efficiently realize the above application, several challenges are identified. The first one is that different video formats produced by the front-end cameras may complicate the video management and analysis task. Besides, the large size of the "raw" video brings severe pressure on the storage disks and processing power. Finally, it is obligatory that the videos redistributed to the users are compatible with and have the same quality on the heterogeneous terminal devices.

The video transcoding system (VTS) is widely believed to address these challenges. It takes large amounts of video files with various formats as inputs and videos with the uniform format and much less size as outputs. Single-machine-based approach and cloud-based approach are two

options to implement the VTS. Nevertheless, the cloud-based approach has several advantages over the single-machine-based approach, especially for the WSN application scenario. Scalability and fault tolerance are two major advantages, for instance. Scalability means that Hadoop nodes are able to easily join/leave the WSN if more/less video streams require transcoding. The VTS continues operating well even if some Hadoop-nodes fail to work, which is the frequently happening case for the WSN applications.

This paper aims to introduce a cloud-based, more specifically, Hadoop-based, video transcoding system to fulfill the vision of bearing hundreds of HD video streams in the next generation WSN, with a discussion on optimization of several significant parameters.

We carry out two sets of experiments: (1) various numbers of Mappers and (2) different file sizes with fixed video duration. Our results show that selecting a proper number of Mappers with respect to the size of video file will obtain optimal performance in terms of transcoding time. The transcoding time depends principally on the duration of video files rather than their sizes.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 discusses the whole VTS architecture and function components. In Section 4, we share our experiment design and result analysis. Section 5 concludes this paper and points out some future works.

2. Related Work

In recent years, the research area on video transcoding has gained more and more attention. A Hadoop-based distributed video transcoding system [2] is designed by using HDFS [3] (Hadoop Distributed File System) to store video resources and applying both MapReduce [4] and FFMPEG [5] technology to do transcoding. A video segmentation strategy on distributed file system (i.e., HDFS) is proposed, in which performance test is also conducted and the result shows that the 32MB size of segment has the most outstanding performance. However, there are still many works left untouched in the paper, such as other options of Hadoop parameters to optimize, for example, block size, cluster size, and so forth. In another paper [6], cloud-based smart video transcoding system is proposed. A cloud server is a logical server that is built, hosted, and delivered through a cloud computing platform over the Internet. It is responsible for providing and managing the user's applications, such as VOD on mobile phones, and storing video materials in the cloud. Service subscribers may access the computing resources by their own computers, smart phones, and so forth. The cloud computing technology could speed up video transcoding, owing to the benefits of parallel computing and MapReduce, that is, splitting original big video file to small portion and forwarding to different nodes in cloud environment and executing transcoding of each small portion in parallel. However, in this paper, no transcoding performance was evaluated. Similarly, another Hadoop-based distributed video transcoding system in a cloud computing environment is proposed to transcode various video codec formats into the MPEG-4 video format [7]. In this system, MapReduce framework, HDFS (Hadoop Distributed File System) platform, and media processing library Xuggler are applied to implement it. Meanwhile, performance evaluation in a Hadoop-based distributed video transcoding system is also conducted [8]. In order to present optimal Hadoop options for processing video transcoding, the experiment data is collected with changing cluster size, block size, and block replication factor. However, these experiments are not enough to figure out the relationship between Hadoop and video transcoding, for example, how to set the number of Mappers for certain video files.

In this paper, we introduce a VTS with Apache Hadoop [9] (including MapReduce framework and HDFS platform, etc.), media processing library FFMPEG, and web server Apache Tomcat [10], in order to verify how it could speed up video transcoding for big video files. This paper also analyzes the experiment results with consideration of factors about the number of Mappers that do the transcoding work, and duration and size of video files, which have not been studied before. It fills the gap in the research field of Hadoop-based video transcoding and contributes greatly to the choice of optimal parameters.

3. Application Scenario and System Architecture

3.1. Application Scenario. Figure 1 illustrates the application scenario of the Hadoop-based VTS in the WSN.

Video capture devices, such as cameras, generate large amounts of videos and transfer them to Hadoop subclusters through the WSN. Then the video transcoding system deployed on the Hadoop cluster will transcode these video files into the MP4 files and store the transcoded files in the HDFS. The system clients, who may use smart watches, smart phones, and other wearable devices, are able to access the system web server to request for transcoded videos or receive recommended videos from the Hadoop cluster. They can also act as video content providers and then submit their own videos to launch new transcoding tasks.

There may be several Hadoop subclusters in one WSN. These Hadoop subclusters can be deployed on the same WSN node or different nodes. From a logical view point, all of these subclusters constitute the whole Hadoop cluster of one VTS. The decentralization property of Hadoop system matches quite well with that of WSN. Seamlessly integrating the Hadoop with WSN would make the solution more robust, scalable, and fault-tolerant.

3.2. System Architecture. This subsection describes the architecture of the video transcoding system. As shown in Figure 2, it consists of a web server and a Hadoop cluster. The web server accepts users' requests and invokes remote video transcoding function deployed on Hadoop cluster. In the following paragraphs, we will introduce each part of the video transcoding system.

(1) Hadoop Cluster. Hadoop is an open-source software framework that derives from Google's Map/Reduce and

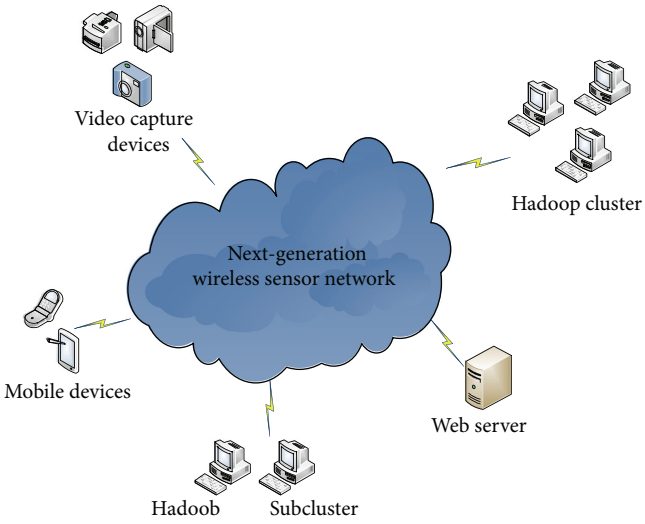


Figure 1: Application scenario.

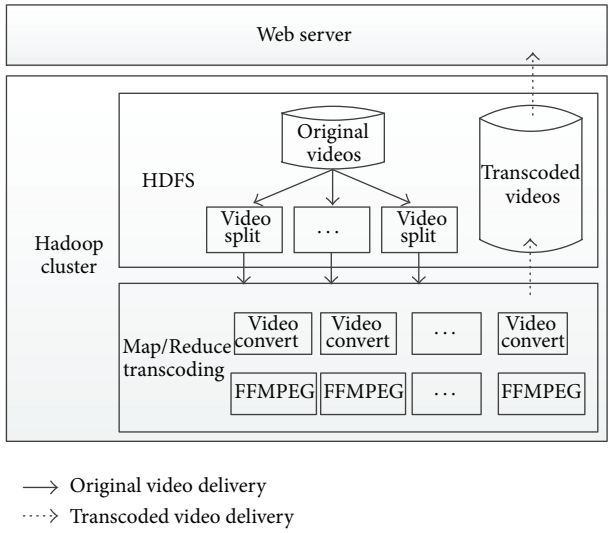


Figure 2: System architecture.

Google File System papers. It provides strong distributed computing capability and scalable storage capacity for huge data processing based on the Hadoop cluster composed by a set of commercial and low cost machines. In the Hadoop cluster as shown in Figure 3, it incorporates a master/slave model; that is, a Hadoop cluster usually includes a single machine designated as a master node and all others act as slave nodes. Based on this model, Hadoop provides a distributed file system known as HDFS and the Map/Reduce function, which are used for data storage and data processing.

(2) HDFS. HDFS is a distributed file system implemented in the Hadoop cluster. It stores large data files across dozens of machines. The master node, also known as the NameNode, is the centerpiece of HDFS. It keeps a directory tree for all the files in HDFS and tracks the location of each part of data files. Moreover, it divides large computation tasks, files into small ones, and distributes them to slave node, known as DataNode. The slave nodes provide storage space and computing resources for HDFS. Large data files are split into several parts and stored across slave nodes. And computation tasks are executed on slave nodes. The results are collected by the master node and combined into an output form.

In our system, the HDFS is mounted on the web server using FUSE tool. From a logical view, it can be treated the same as the local file system. The original videos and transcoded videos are stored in HDFS. The NameNode and DataNodes can colocate with one WSN node or connect with different WSN nodes.

(3) Map/Reduce Model. Map/Reduce is the programming model for processing parallelizable problems across large datasets with some distributed algorithms running on a cluster of machines. It involves a Map procedure and a Reduce procedure. As indicated in Figure 4, the input data of Map/Reduce model will be split into small datasets. Then these small datasets are mapped into some Map processors' work space and handled. A shuffle/sorting mechanism will collect results of the Map procedure and organize them into lists. The Reduce procedure will run on these lists and transform the results into the final output data.

As a popular implementation of Map/Reduce model, Hadoop uses the master node to complete the Map and Reduce procedure. In the Map procedure, the master node divides the input, such as large data files or computation tasks, into small ones and then generates (key, value) pairs. Depending on the shuffle and sorting mechanism, it distributes them to the slave nodes. After small tasks are finished, the master node collects results from slave nodes to accomplish the Reduce procedure.

(4) Video Transcoding. The format of a video includes many parameters, for example, bit rate, frame rate, spatial resolution, coding syntax, content, and so forth. Video transcoding is the technology used to convert a video from one format to another one. There are numerous video transcoding mechanisms. Some main transcoding techniques are briefly introduced as follows [11–13].

(a) Bit-rate transcoding: it is usually applied to reduce the bit rate with the same complexity and quality if possible. The likely scenario is to convert the video resources for television broadcast and Internet streaming. There are some bit-rate transcoding architectures: open-loop transcoders, cascaded pixel-domain transcoders, and DCT-domain transcoders.

(b) Spatial and temporal transcoding: apart from bit-rate transcoding, spatial and temporal transcoding is also used to convert compressed video for communication networks video viewer. There are a multitude of challenges, one of which is how to derive a new set of motion vectors. Many relevant researches have been performed to solve them. In addition, some architectures are proposed to enhance its performance, such as DCT-domain architecture, a hybrid DCT/pixel-domain transcoder architecture.

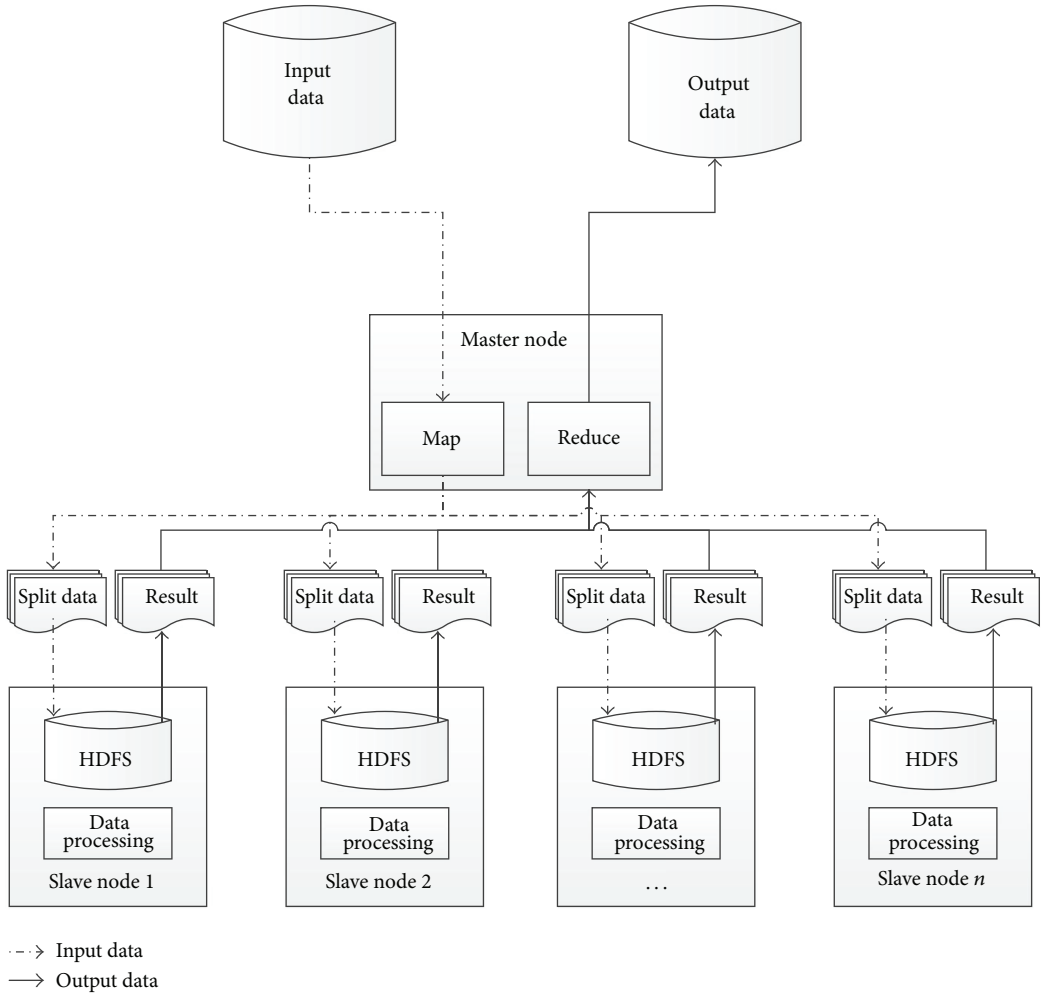


Figure 3: Hadoop cluster work flow.

(c) Standards transcoding: coding standards, for example, MPEG-2, MPEG-4, and so forth, are one of the important characteristics of video. Besides the conversion in bit rate and resolution, coding standard will also be changed from one to another in some applications.

Our video transcoding mechanism, by itself, is based on the spatial and temporal transcoding technique together with standards transcoding technique. We conduct video transcoding from AVI to MPEG-4 and meanwhile from 800×480 to 640×240. More specifically, we use a Mapper to divide large video files into small ones. Afterwards, these small files are distributed to be stored in HDFS. There is no Reduce procedure in order to eliminate the time consumption. All of the transcoding work is done on the slave nodes. We use the open-source software FFMPEG to split videos and do transcoding work. FFMPEG is a free multimedia software framework used to handle various multimedia data. It also provides some library files related to audio/video codec. In addition, it can be compiled under most operating systems.

4. Experiments

4.1. Experiment Settings. Figure 5 illustrates the experiment platform deployed on three Dell servers and one PC. The Hadoop cluster is composed of three Dell servers, which serve as one NameNode and two DataNodes. It does not hold lots of meanings if tens of DataNodes are deployed considering that the most usual application locations of WSN are outdoor, in mobile status, or even in barren areas. Each Hadoop cluster node is running on the Linux OS (Debian 3.2.46 x86_64). No matter the NameNode and DataNode are, each Hadoop-node is equipped with two Intel Xeon 8 core 2.00 GHz processors with 64 GB registered ECC DDR memory and 3 TB SATA-2. The web application is deployed on Tomcat web engine running on a PC machine, which is equipped with Linux OS (Ubuntu Server 3.5.0 x86_64), Intel Core 2 Duo CPU 3.0 GHz with 2 GB registered DDR memory, 320 GB SATA-2. Java 1.6.0_37, Apache Tomcat 7.0.39, Hadoop-0.20.2, and FFMPEG 1.0.6 are the other components used in the platform.

Several video data sets are used for the system performance evaluation. The video data sets are generated by emerging replications of the original file using the Format

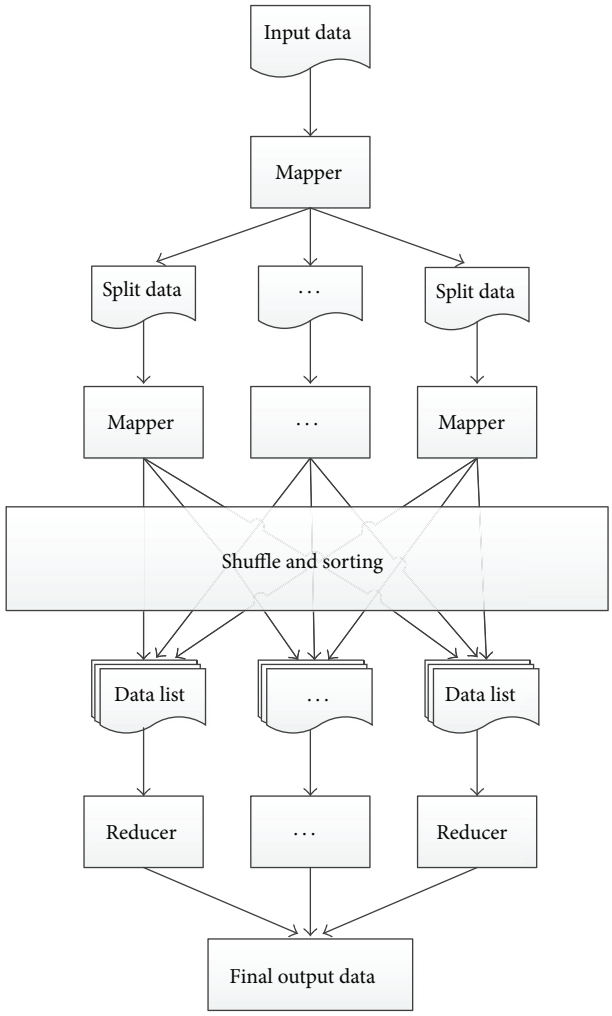


Figure 4: Map/Reduce model.

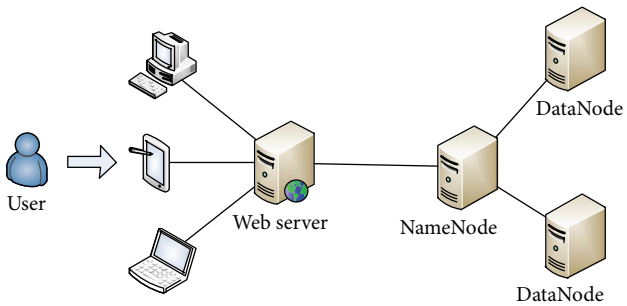


Figure 5: Experimental platform.

Factory [14]. Each original file size is 80MB. Table 1 lists the parameters for original and transcoded video file.

4.2. Experimental Results and Analysis. This subsection focuses on discussing and analyzing the experiment results. The transcoding time consumption is employed as the performance metric. In the experiments, we choose parameters

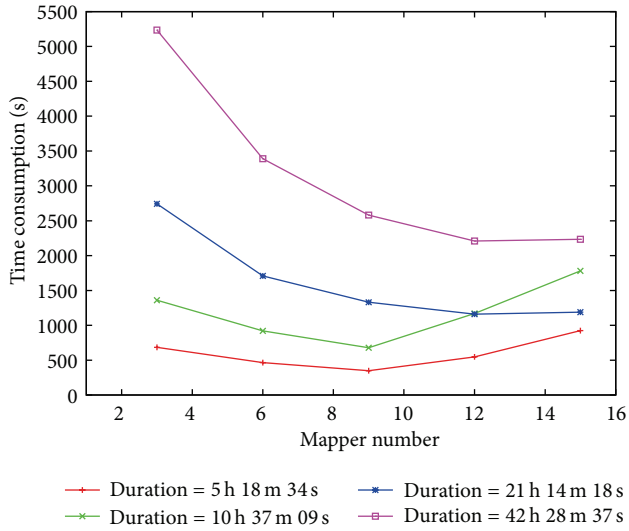


Figure 6: Effect of number of Mappers.

Table 1: Experiment settings.

Parameter	Original single file	Transcoded file
Codec	DivX	MPEG-4
Format	AVI	MP4
Duration	31 min 21 sec	31 min 21 sec
Resolution	800 × 480	640 × 240
Frame rate	15 fps	15 fps

about the number of Mappers that do transcoding work and duration and size of video files. We carry out a series of experiments based on these factors.

(1) Effect of Number of Mappers. We design the first set of experiments by changing the number of Mappers while fixing file size and duration. The range of Mappers' number is 3, 6, 9, 12, and 15. Table 2 shows the video data sets used in these experiments and Table 3 lists the experiment results.

Figure 6 indicates the effect of number of Mappers. For the files whose durations are 5 h 18 m 34 s and 10 h 37 m 9s, as the Mappers' number increases, the time consumptions of transcoding decrease at the beginning and reach the lowest point when the number is 9. Later, they increase obviously as the number of Mappers is greater than 9. For the files whose durations are 21 h 14 h 18 s and 42 h 28 m 37 s, time consumptions go down when Mappers' number increases.

In the Hadoop system, a Mapper is designated to process a split part of video file and is invoked by a DataNode. The more the Mappers, the more the transcoding tasks distributed to DataNode, which accelerates the transcoding process. On the other hand, there are also queues in each DataNode keeping the disengaged transcoding tasks waiting, which retards the transcoding process. There should be a "Laffer curve" just as the case in economics [15].

The curve depicted in Figure 6 reveals that there exists an optimal value of the number of Mappers for the Hadoop-based VTS and the value is closely related to the size of files.

Table 2: Video data sets for performance evaluation.

File size	1 GB	2 GB	4 GB	8 GB
Total duration	5 h 18 m 34 s	10 h 37 m 09 s	21 h 14 m 18 s	42 h 28 m 37 s

Table 3: Transcoding time with different Mappers' numbers.

Mapper	Video data sets			
	1 GB/5 h 18 m 34 s	2 GB/10 h 37 m 09 s	4 GB/21 h 14 m 18 s	8 GB/42 h 28 m 37 s
3	684 s	1359 s	2743 s	5235 s
6	464 s	921 s	1709 s	3390 s
9	349 s	678 s	1330 s	2581 s
12	547 s	1170 s	1160 s	2210 s
15	925 s	1781 s	1189 s	2234 s

Table 4: File sizes and results—first file.

First file size	Time consumption
0.7 GB	317 s
1.0 GB	349 s
1.3 GB	309 s
2.0 GB	340 s

Table 5: File sizes and results—second file.

Second file size	Time consumption
1.7 GB	578 s
2.0 GB	678 s
2.9 GB	576 s
4.0 GB	601 s

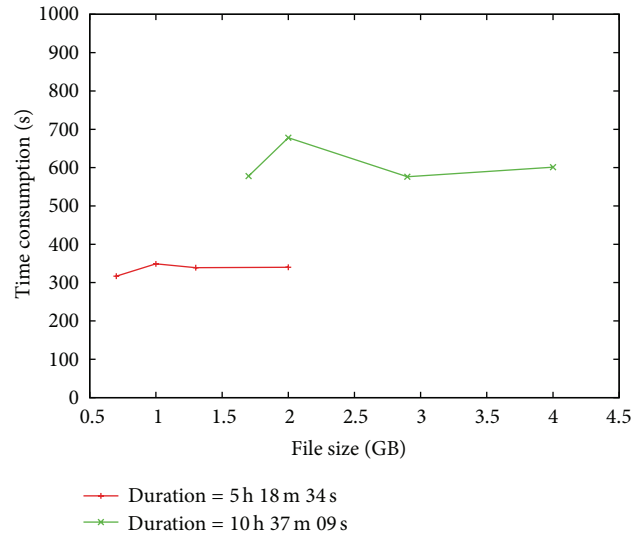


Figure 7: Effect of duration and size.

The optimal value is 9 for the video files whose durations are 5 h 18 m 34 s and 10 h 37 m 9s in experiment settings of this paper, while, for the other two cases, the optimal value should be around 12.

(2) Effect of Duration and Size. The second set of experiments tries to explore the relationship of the transcoding time with the size and duration of video files. In these experiments, we specify the Mappers' number as 9. By fixing the duration of two video files and changing their sizes, we investigate the transcoding time consumption. Tables 4 and 5 indicate the system performance with respect to various file sizes. The durations of two video files are 5 h 18 m 34 s (Table 4) and 10 h 37 m 09 s (Table 5), respectively.

Figure 7 demonstrates that time consumptions fluctuate a little when file sizes increase and meanwhile the durations are fixed. The time consumption of video transcoding depends principally on the duration of video files rather than their sizes. It suggests that duration-based splitting mechanism would be more controllable than the size-based method.

5. Conclusion and Future Work

In this paper, we propose a Hadoop-based VTS integrating several key components including HDFS, Map/Reduce, FFMPEG, and Tomcat, with a discussion on several significant parameters. Three prominent results are achieved through the experiments: (1) it is clear that there is an optimal value of the number of Mappers; (2) the optimal value is closely related to the file size; (3) the time consumption of video transcoding depends principally on the duration of video files rather than their sizes.

The inherited distribution property of the Hadoop system seems quite harmonious with the decentralization attribute of next-generation WSN. However, the research exploring their relationship and integrating them into a turn-key solution for many practical problems is still in its preliminary stage. As one of the first papers pioneering in this direction, this paper enlightens several directions for future work.

First of all, more experiments will be carried out not only in the area of standards transcoding and spatial transcoding, but also in the field of bit-rate transcoding, to meet the service requirements of the next-generation WSN, such as converting the video resources for video broadcast or streaming.

Secondly, some efforts have to focus on WSN's network planning and routing protocol optimization in order

to integrate seamlessly with Hadoop system. The normal Hadoop system generally operates in an indoor and machine-friendly environment with wired connections. However, the WSN system always works in outdoor locations with tough surroundings, such as severe interferences or extremely high or low temperatures, and so forth. Accordingly, the communication protocols among Hadoop nodes also have to be redesigned with the consideration of WSN characteristics.

Besides, some theoretical research will be done to find the optimal value of the Mappers' number. The mathematical model is going to be constructed, taking the Hadoop cluster size, block size, video file size, and block replication factor into account.

Finally, it is necessary to implement the Hadoop-based online or real time VTS for the next generation WSN, beside the offline version proposed in this paper. It makes the live video broadcasting, multicasting, and P2P streaming possible in the WSN with the online VTS at hand.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is supported by Science and Technology Commission of Shanghai Municipality Research Project "Research of key technologies and applications of FOD detection in aircraft movement area" (Project nos. 13511503200 and 13511503202).

References

[1] O. Bejarano, E. W. Knightly, and M. Park, "IEEE 802.11ac: from channelizaiton to multi-user MIMO," *Communications Magazine, IEEE*, vol. 51, pp. 84–90, 2013.

[2] F. Yang and Q. W. Shen, "Distributed video transcoding on Hadoop," *Computer Systems & Applications*, vol. 20, no. 11, pp. 80–85, 2011.

[3] HDFS, http://en.wikipedia.org/wiki/HDFS#Hadoop_distributed_file_system.

[4] MapReduce, http://en.wikipedia.org/wiki/Mapreduc.

[5] FFMPEG, http://www.ffmpeg.org/.

[6] A. A. Chandio, I. A. Korejo, Z. U. A. Khuhro et al., "Clouds based smart video transcoding system," *Sindh University Research Journal*, vol. 45, pp. 123–130, 2013.

[7] M. Kim, Y. Cui, K. Lee et al., "Towards efficient design and implementation of a Hadoop-based distributed transcoding system," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 2, pp. 213–224, 2013.

[8] N. S. Chahal and B. S. Khehra, "Performance evaluation of a Hadoop-based distributed video transcoding system for mobile media service," ASTL Volume 2, Information Science and Technology(Part 1), 2012.

[9] The Apache Hadoop Project, http://hadoop.apache.org/.

[10] The Apache Tomcat Project, http://tomcat.apache.org/.

[11] Y. Wu, A. Vetro, H. Sun, and S. Y. Kung, "Intelligent multi-hop video communications," in *Presented at the IEEE Pacific Rim Conference on Multimedia*, Beijing, China, 2001.

[12] S.-F. Chang and A. Vetro, "Video adaptation: concepts, technologies, and open issues," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 148–158, 2005.

[13] J. Xin, C.-W. Lin, and M.-T. Sun, "Digital video transcoding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 84–97, 2005.

[14] Format Factory, http://www.pcfreetime.com/.

[15] Laffer curve, http://en.wikipedia.org/wiki/Laffer_curve.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 269596, 8 pages
http://dx.doi.org/10.1155/2014/269596

Research Article

RFID Localization Using Angle of Arrival Cluster Forming

Waleed Alsalih,[1] Abdallah Alma'aitah,[2] and Wadha Alkhater[1]

[1] *Department of Computer Science, College of Computer and Information Sciences, King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia*
[2] *School of Computing, Queen's University, Kingston, ON, Canada*

Correspondence should be addressed to Waleed Alsalih; wsalih@ksu.edu.sa

Received 29 November 2013; Accepted 15 January 2014; Published 20 March 2014

Academic Editor: Christos Verikoukis

Radio Frequency IDentification (RFID) has been increasingly used to identify and track objects automatically. RFID has also been used to localize tagged objects. Several RFID localization schemes have been proposed in the literature; some of these schemes estimate the distance between the tag and the reader using the Received Signal Strength Index (RSSI). From a theoretical point of view, RSSI is an excellent approach to estimate the distance between a sender and a receiver. However, our experiments show that there are many factors that influence the RSSI value substantially and that, in turn, has a negative effect on the accuracy of the estimated distance. Another approach that has been recently proposed is utilizing transmission power control from the reader side. Our experiments show that power control results are more stable and accurate than RSSI results. In this paper, we present a test-bed comparison between the power control and the RSSI distance estimation approaches for active RFID tags. We also present the Angle of arrival Cluster Forming (ACF) localization scheme that uses both the angle of arrival of the tag's signal and the reader's transmission power control to localize active tags. Our experiments show that ACF is very accurate in estimating the location of active RFID tags.

1. Introduction

Radio Frequency IDentification (RFID) is a wireless technology that has enabled a wide range of identification, monitoring, and tracking applications. This includes using RFID for supply chain management, access control, identification and tracking of patients and children, and smart environments.

An RFID system is composed of tags (or transponders), readers, and a software application. A tag is usually attached to an item for the purpose of identifying and/or tracking it automatically. RFID tags store some identification data and transmit this data to the RFID reader via Radio Frequency (RF) signals. The RFID reader broadcasts a query to all tags within its reading (interrogation) range. Tags that receive the broadcasted query respond by sending their data back to the reader [1, 2].

RFID tags are composed of three main components: an Integrated Circuit (IC), an antenna, and substrate. Based on the power source, tags can be classified into three types: passive tags, semipassive tags, and active tags. Passive tags do not have a power source; they rather get powered by the electromagnetic waves that are emitted from the reader to broadcast the query [1–3]. Semipassive tags have an onboard battery that is used only to power their circuits for processing. However, the battery is not used for communication. As is the case with passive tags, semipassive tags rely on the reader to get the power they need for communication. An active tag has a battery that is used for both powering up the circuit and communicating with the reader. An active tag can handle a two-way communication with the reader independently [4].

RFID has the potential to enable many interesting applications and to play a crucial role in the envisioned Internet of Things (IoT). However, achieving that requires solving some technical problems. One of these problems is estimating the location of a tag precisely, which is the focus of this paper. Accurate localization of an RFID tag is the key to many tracking and monitoring applications. An example of such applications is an alarm system that predicts dangerous situations (e.g., a child walking toward an oven).

Several RFID localization schemes have been proposed in the literature. Most of these schemes find the location of a tag based on an estimate to the distance between the tag and multiple readers. The accuracy of the estimated location depends primarily on the estimated distances to multiple readers.

Estimating the distance between a reader and a tag can be made either by using the Received Signal Strength Index (RSSI) or by controlling the transmission power of the reader.

In this paper, we look into the problem of localizing an active RFID tag precisely. We have conducted a test-bed experiment to evaluate and compare RSSI-based and power control-based schemes. We also propose the Angle of arrival Cluster Forming (ACF) scheme which uses a combination of the Angle of Arrival (AoA) and the reader's transmission power control. Our experiments have shown that ACF is very accurate in estimating the location of active RFID tags.

The rest of the paper is organized as follows. Section 2 surveys related work. Section 3 presents our test-bed experiments for RSSI-based and power-control-based distance estimation approaches. Section 4 presents our ACF scheme for active RFID localization. Section 5 concludes the paper with a brief summary and some future research directions.

2. Related Work

Many localization schemes have been proposed in the literature. The SpotON system, which was proposed in [8], is a general wireless localization scheme that uses RSSI measurements and the triangulation method. The scheme is based on using multiple base stations (readers) that provide RSSI measurements for the target object (tag) to be used to estimate its location. However, this system has a low accuracy in RFID because of the unpredictable behavior of the RF signals especially with passive tags.

A famous RFID localization scheme is the LocAtioN iDentification based on dynaMic Active Rfid Calibration (LANDMARC) which was proposed to localize active tags in indoor environments [9]. LANDMARC uses reference tags to serve as reference points in known locations. LANDMARC estimates the location of a target tag whose location is unknown by comparing its RSSI with those of reference tags. LANDMARC is considered to be a cost-effective localization scheme as it uses more reference tags instead of using more readers. The authors in [10] proposed the VIrtual Reference Elimination (VIRE) scheme which extends the LANDMARC scheme by using virtual reference tags in addition to the actual reference tags. RSSIs of virtual reference tags are calculated and used to enhance the accuracy of the localization process. Another enhancement to the LANDMARC scheme was presented in [11]. This scheme divides the area into a number of subareas and finds two estimates for the location. The first estimate is calculated using the LANDMARC scheme and the second estimate is calculated using a new estimation algorithm. The estimation algorithm starts by creating a vector of average Euclidean distances between each area and the target object. The area with the lowest average Euclidean distance is chosen to be the second location estimate for the target object.

In [12], a scheme for localizing active and passive mobile tags is proposed. It uses RSSI and reference tags placed on the floor and the ceiling of the environment. In [13], the authors proposed an approach that utilizes connectivity information of readers and virtual reference tags only to locate a target tag. Experiments have shown a fine-grained accuracy for 2D localization as well as an acceptable accuracy for 3D localization. The authors in [14] presented an RFID smart shelf using passive UHF RFID tags to localize tags by detecting interference with reference tags.

Another approach that can be used is utilizing power control from the reader side which was proposed initially as an anticollision protocol [15–17]. The main idea is to adjust the reader's transmission power level (i.e., its interrogation range) in order to estimate the distance between the reader and the tag. The distance is associated with the lowest power level at which the tag is detected. In other words, the power level is mapped to a distance [18].

In the literature, several AoA-based schemes are proposed for localizing UHF RFID tags. In [6, 7], three uniform linear antenna arrays with three patch elements are used to localize active tags using the AoA. The AoA of the received tag signal is measured at each of the three antenna array positions. However, previous work did not pay attention to errors in AoA measurements and their effect on the localization process. The work in [19–21], for example, assumes that AoA measurement error is uniform over all angle readings.

More details about existing localization schemes can be found in some survey papers about this topic (see [22, 23]).

In this paper, we examine the sole use of RSSI or power control approaches as an initial, yet coarse, step to localize active RFID tags. Even though AoA is an attractive localization method, we show that existing AoA estimation methods comprise high error rates due to the phase measurement correlation. Therefore, we propose a localization method that augments the power control approach with a new AoA method that minimizes the phase measurement correlation and, hence, provides higher accuracy.

3. Test-Bed Evaluation of RSSI and Power Control Distance Estimation

Since most existing distance estimation schemes are based on either tags RSSI or readers power control, we present a test-bed evaluation for both approaches. We use the following devices in our test-bed. We use GAO RFID 217001B active reader and GAO 127002 active tags [24]. The interrogation range of the reader can be adjusted from 5 m to 100 m, and it can read up to 100 tags/second. Its operating frequency ranges from 2.4 GHz to 2.5 GHz. This reader comes with an Application Programming Interface (API) library that makes an interface for the reader to control its operation by external software. It has a dynamic transmission power that is configurable by an API function.

3.1. RSSI Distance Estimation. We evaluated the approach of using RSSI (only) for estimating the distance between readers and tags. The distance is estimated based on mapping the RSSI readings of the tag's response to a distance based on a reference curve (i.e., a mapping table) associating RSSI readings to distances. We placed tags at distances that range from 1.5 m to 15 m with intervals of 1.5 m (i.e., 10 locations)

with 10 dBm transmission power from the reader. We had the following observations.

(1) The same tag at the same location gives different RSSI readings.

(2) A tag may give an RSSI reading that is larger than that of another tag that is closer to the reader.

The reason behind these two observations is the fact that RSSI in general is unstable and very sensitive to tags orientation and different environment effects. We also took the average RSSI reading out of multiple readings with different orientations. The result is illustrated in Figure 1 which shows the average RSSI reading and the variance for each distance. The obtained results are unstable and inaccurate. For instance, at short distances, the readings of a tag placed at 4.5 m from the reader overlap with the readings of a tag placed in the range from 4.5 m to 9 m, which produces an error margin of 4.5 m (i.e., 100% of the actual distance). At longer distances, the readings of a tag placed at 10.5 m overlap with the readings of a tag placed in the range from 10.5 m to more than 15 m. This makes RSSI-based localization unsuitable for both indoor and outdoor applications.

RSSI fluctuations are worse with passive tags because of the use of backscatter modulation [4] to reflect the reader's low-power signal. In [5], an evaluation of RSSI efficiency for distance estimation is conducted and the results are shown in Figure 2. Figure 2 shows the distribution of the measured RSSI and the mean of these values under the following conditions: 30 dBm transmission power, 1 m height for antennas and tags, and 6 dBi reader's antenna gain. The reported results exhibit a similar error trend as that of the active tags in Figure 1.

The main conclusion we have out of this experiment is that RSSI cannot be used to accurately measure the distance between a tag and a reader. In fact, even reference tags do not help here as it is common to have very different RSSI readings from the same tag in the same location and at almost the same time.

3.2. Power Control Distance Estimation. Another method to estimate the distance between a tag and a reader is to adjust the transmission power and, hence, the transmission range of the reader. This way the distance is associated with the minimum transmission range at which the tag is identified as shown in Figure 3. We conducted a comprehensive experiment to evaluate the stability and accuracy of this method. We placed tags at distances that range from 1 m to 19 m with intervals of 2 m (i.e., 10 locations). The RFID reader has 31 transmission levels. In our experiment, the reader interrogates tags 100 times at each transmission level. For each combination of a distance and a transmission level, we find the percentage of times the tag is identified by the reader. The results of this experiment are shown in Table 1. The results are much more stable than the RSSI results. We observed that tags located in the same location give very consistent results.

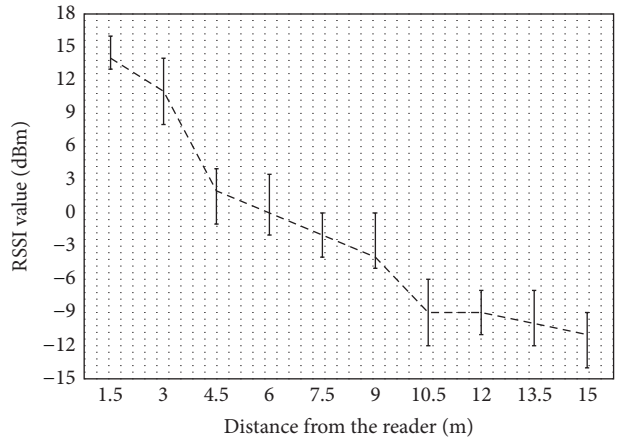


Figure 1: RSSI values of active tags from an active reader with 10 dBm transmission power.

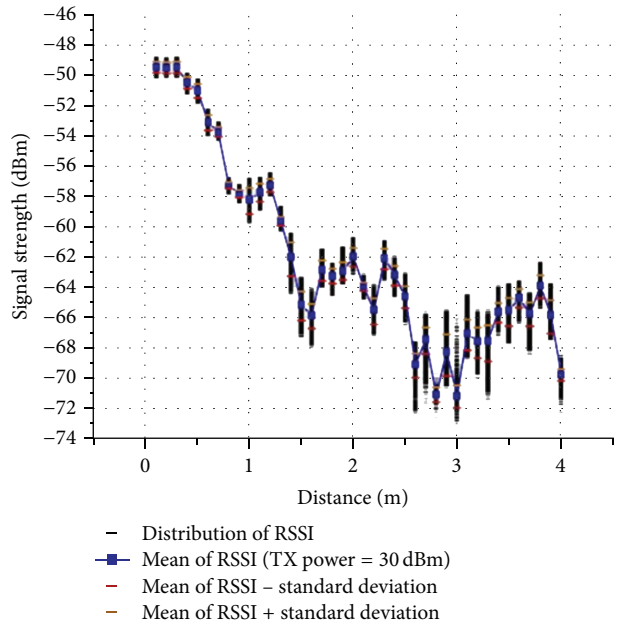


Figure 2: Relationship between RSSI values and distance for passive tags [5].

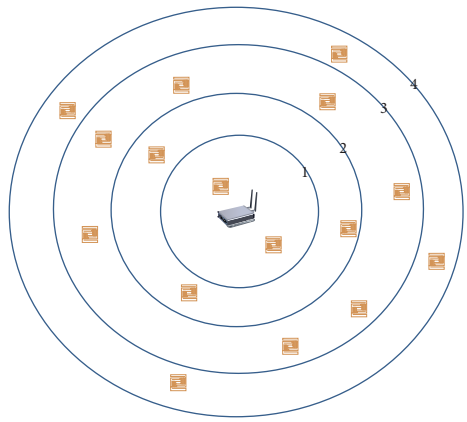


Figure 3: Estimating the distance using power control.

Table 1: Relationship between distance and transmission power.

		Power level															
		1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31
Distance (m)	1	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	3	0	0	0	52	98	98	98	99	99	100	100	100	100	100	100	100
	5	0	0	0	38	49	50	49	49	49	50	50	50	51	49	51	49
	7	0	0	0	16	29	34	49	51	50	51	42	47	46	48	49	49
	9	0	0	0	0	0	22	32	49	49	54	50	47	54	49	48	51
	11	0	0	0	0	0	6	32	23	41	48	40	43	50	48	50	52
	13	0	0	0	0	0	0	21	21	19	20	27	45	49	51	53	50
	15	0	0	0	0	0	0	12	13	14	12	19	36	37	49	44	51
	17	0	0	0	0	0	0	1	10	10	7	13	25	28	41	42	51
	19	0	0	0	0	0	0	0	3	4	6	7	17	24	34	34	51

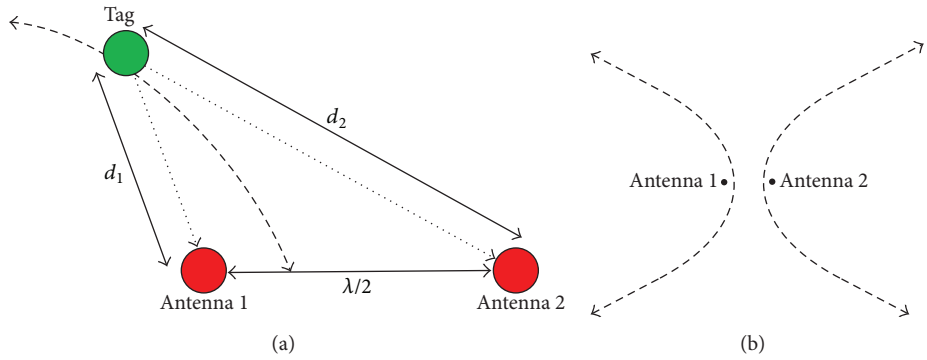


Figure 4: (a) Phase difference calculation. (b) The hyperbolas of antenna 1 and antenna 2.

4. Angle of Arrival Cluster Forming (ACF)

In the previous section, we have seen that power control can give good estimates for the distance between a reader and a tag. This divides the area surrounding the readers into discrete regions in which tags are expected to reside. In this section, we propose a scheme that augments the highly precise AoA localization method with the power control distance estimation method. We implement a method for estimating the AoA using the phase difference in the 2.45 GHz range. Experimental results are presented to assess the performance of the proposed scheme.

4.1. Overview. Finding the AoA of the tag's signal is done by estimating the phase difference of the received tag's signal with respect to known positions of the receiving antenna arrays. And then, the position of the tag is determined to be in the intersection of the two hyperbolas formed by the two antenna arrays.

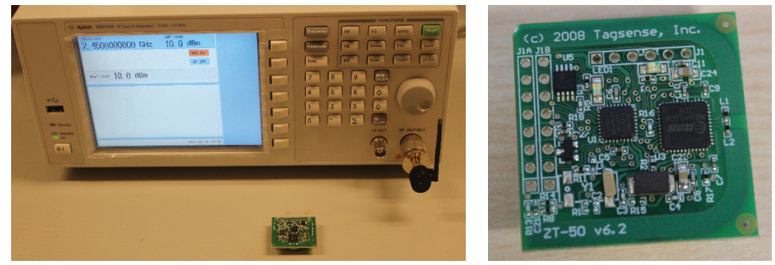
To estimate the angle of arrival of the tag's signal, we use a phased antenna array composed of two individual antenna elements. The two antenna elements are separated by a distance of $\lambda/2$ m (where λ is the frequency wavelength) and aligned along a horizontal line to form a uniform linear array (see Figure 4). The tag's signal arrives to the two antennas at different times, which makes a constant phase shift. The phase detector chip translates that phase difference to a specific voltage. The phase difference (θ_d) is equal to $360(d_1 - d_2)/\lambda$, where d_1 and d_2 are the distance between the tag and the first antenna in the array and the distance between the tag and the second antenna in the array, respectively.

The two-antenna array provides a solution for the tag's location which is a hyperbola whose focal points are the locations of the two antennas (as shown in Figure 4(b)). If the tag is placed at a point (x_t, y_t), the phase difference is θ_d, and the two antennas are placed at two points $(-\lambda/4, 0)$ and $(\lambda/4, 0)$; the solution hyperbola for the tag's location is defined by the following equation:

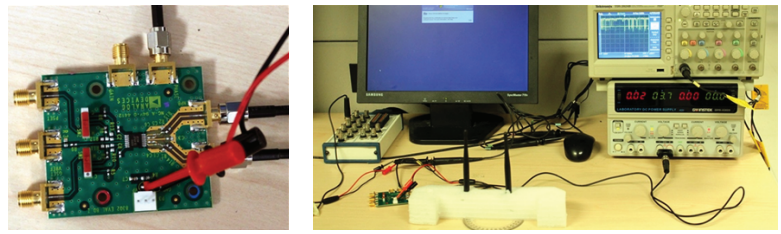
$$y = \sqrt{\frac{x\lambda^2}{16a^2} - x^2 - \frac{\lambda^2}{16} + a^2}, \quad \text{where } a = \frac{\lambda}{4}\left(\frac{\theta_d}{180}\right). \tag{1}$$

This indicates that a single antenna array of two antennas is not enough for a precise localization because any point on the hyperbola is a possible location of the tag. Therefore, deploying more antenna arrays brings overlapping hyperbolas and provides a finer estimate for the tag location.

4.2. Error Modeling. In our design, we utilize the phase difference using the AD8302 chip to estimate the AoA [25]. In order to evaluate the localization accuracy, we conducted an experiment to evaluate the reflected error in the AoA from the error in the measured phase difference. A two-antenna array (with $\lambda/2$ m distance between the omnidirectional antennas)



(a)



(b) (c)

FIGURE 5: (a) RF signal generator and an active tag. (b) AD8302 phase shift detector board. (c) The two-antenna array is connected to the phase difference module and the output is connected to the oscilloscope and data acquisition board.

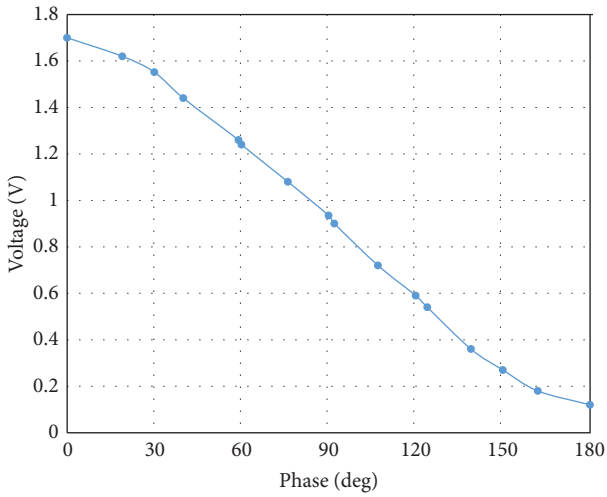


FIGURE 6: The output voltage from the phase difference module when shifting the antenna array from 0 to 90 degrees (causing a phase difference from 0 to 180 degrees).

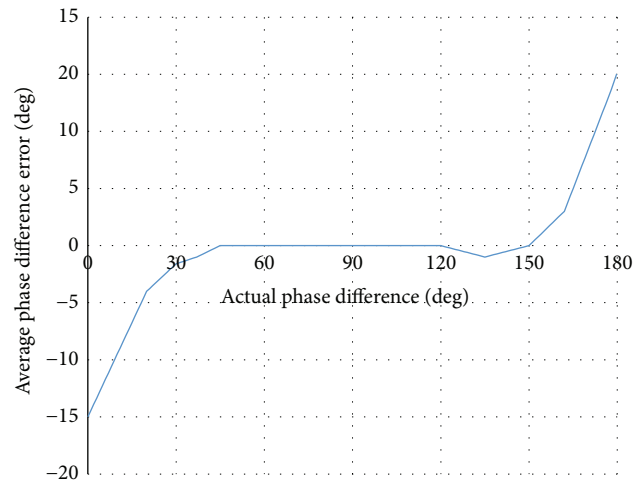


FIGURE 7: The error (in degrees) between the actual and the measured phase shifts.

is placed at a distance $d = 5$ m from the RF signal generator as shown in Figure 5. The output power of the signal generator is set to 10 dBm (excluding the antenna gain) with an output frequency of 2.45 GHz. The antenna array is then rotated from 0 to 90 degrees (around its center point between the antennas) in 5-degree steps. Since the distance between the antennas is $\lambda/2$ m (or 180 degrees), the 90 degrees rotation is reflected in θ_d of 180 degrees. For each step the output voltage is plotted by the oscilloscope and mapped to the phase shift as shown in Figure 6. The difference between the actual and the measured phase shifts is then plotted as shown in Figure 7. Note that the error (in degrees) is not constant for all angles of the phase shift and we consider this fact in the evaluation of our system's performance.

4.3. Deployment of Phase Detectors. When the two detectors are placed beside each other as in [6, 7, 21], the phase estimation error has a significant impact on the location estimation. As shown in Figure 8(a), for each phase detector n there is a phase error margin that results in upper and lower limits to the expected phase, θ_{1_n} and θ_{2_n}. Each phase limit creates a hyperbola, and the tag's location can be anywhere between the two hyperbolas. In Figure 8(a), the tag is assumed to be 4λ m from the phase detectors. The hyperbolas of the two phase detectors intersect in an area spanning more than 4λ m (around 8.5λ m) because of the correlation between the measured phases which is reflected in the adjacent hyperbolas.

To provide a fine estimate for the tag's location, we place two phase detectors at a distance that is much larger than the

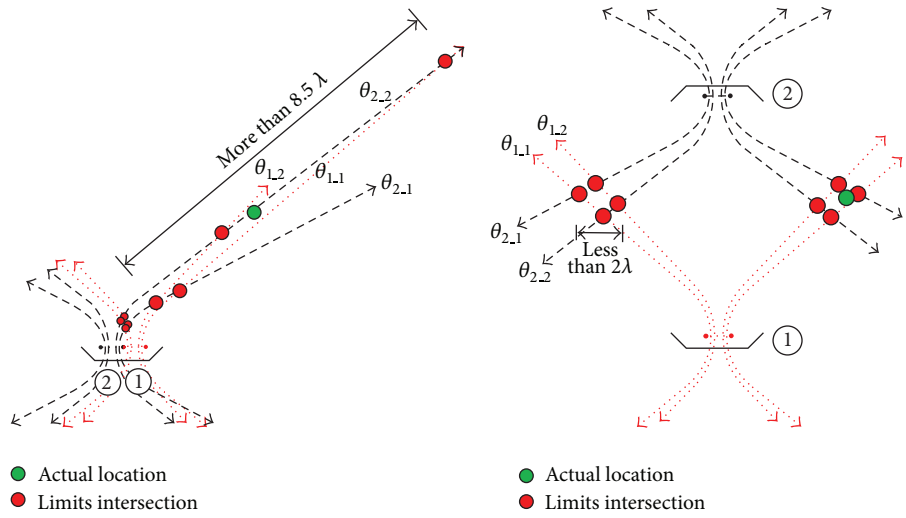


FIGURE 8: The drawback of antenna arrays with adjacent antenna array elements [6, 7]. (b) The advantage of separating the antenna elements in the environment (high precision).

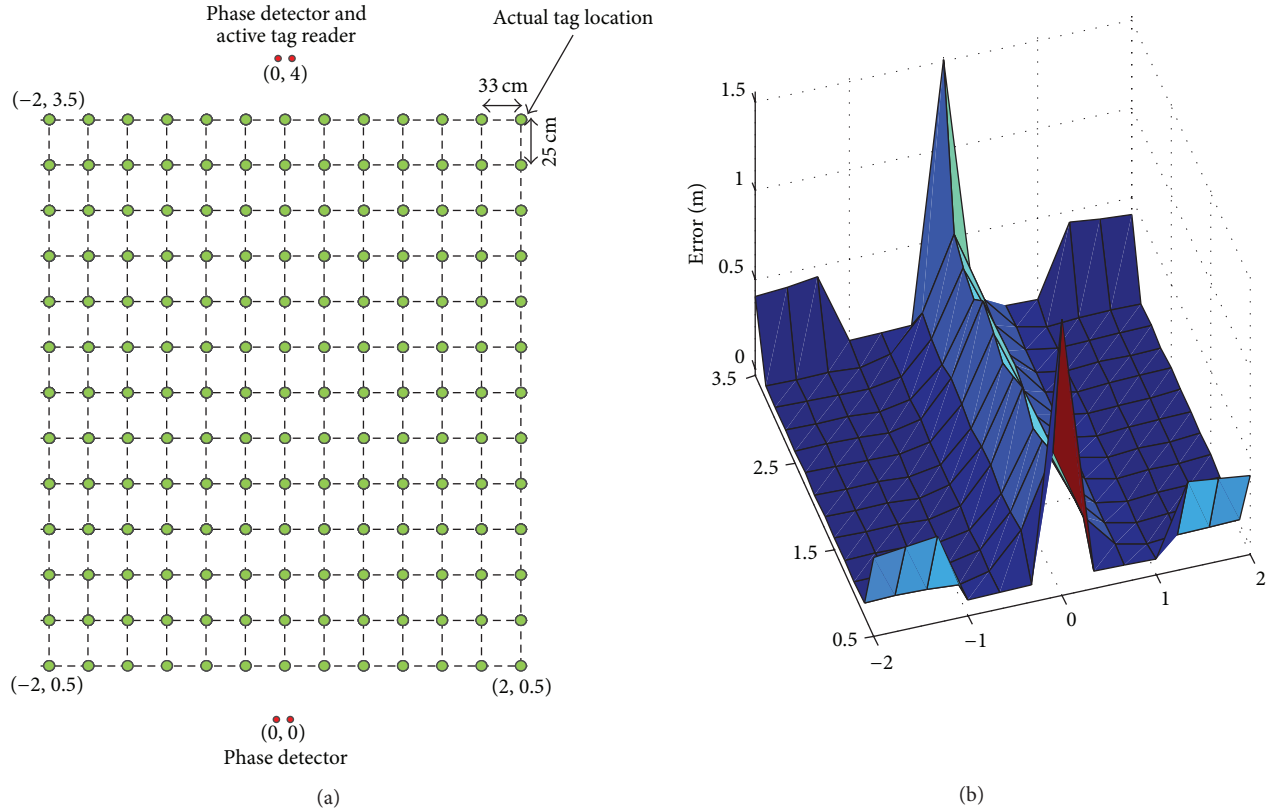


FIGURE 9: (a) The grid in our experiment. (b) Estimated error for each point in the grid in (a).

wavelength λ. The phase reading from each detector makes a hyperbola and the tag's location is estimated to be along the edge of that hyperbola. As the phase reading has a variable margin of error [25], we consider the upper and lower limits of the phase difference in estimating the location. When the phase detectors are separated as shown in Figure 8(b), the same phase error results in a much smaller error area than that in Figure 8(a). The deployment in Figure 8(b) results in two clusters of points which are close to each other.

4.4. Combined Power Control and ACF. The power control scheme provides a good estimate for the distance between the

tag and the reader. Thereby, the ACF can utilize the power control distance estimation to consider only one of the location clusters (as in Figure 8(b)). In our evaluation, two phase detectors and one active tag reader are placed as shown in Figure 9(a) with a separation of 4 m (with an operating frequency of 2.4 GHz). An active tag is placed at the grid nodes in Figure 9(a) that covers an area of $3 \times 4\ m^2$. The experiment area is then mapped to the Cartesian plane with the lower corners at (−2, 0.5) and (2, 0.5) and the upper corners at (−2, 3.5) and (2, 3.5) with one phase detector at the origin and another at (0, 4). The localization error is calculated based on the Euclidian distance between the estimated location and the actual location. The error at each node is plotted in Figure 9(b). Note that most of the nodes have an error of less than 0.1 m indicating a robust localization performance of our scheme. When the tag is at the middle nodes, the phase difference becomes close to zero and the error rises dramatically as depicted in Figures 7 and 9(b). The error rises also when the tag is at the corner of the grid where the phase difference is close to 180 degrees. The average error of the 169 locations is 0.135 m.

5. Conclusion

In this paper, we present a test-bed study and comparison between the RSSI and the power control distance estimation approaches for active RFID tags. We found that the power control approach is much more stable and accurate than the RSSI approach. We also present a novel scheme for localizing active tags, which is the ACF deployment, that takes advantage of the robustness of the AoA of the tag's signal and the accurate estimated distance from the power control approach to localize active tags accurately. Our experiments have shown that a combination of the ACF and power control is able to locate active tags with a high accuracy (an average of 13.5 cm error) using a single RFID reader (with adaptive gain) and two phase detectors.

Our future work includes combining the ACF and power control with the use of reference tags to improve the accuracy in some particular spots in which the AoA module is expected to suffer low accuracy (e.g., middle nodes and corners of the grid).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research is supported by the National Plan for Science and Technology Program at King Saud University, Project no. 11-INF1500-02.

References

[1] M. Brown, S. Patadia, S. Dua, and M. Meyers, *Mike Meyers' Comptia RFID+ Certification Passport*, Mcgraw-hill, 2007.

[2] F. Thornton, D. P. Sanghera, B. Haines et al., *How To Cheat at Deploying and Securing RFID*, Elsevier Science, 2011.

[3] S. Preradovic, N. C. Karmakar, and I. Balbin, "RFID transponders," *IEEE Microwave Magazine*, vol. 9, no. 5, pp. 90–103, 2008.

[4] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication*, Wiley, 2003.

[5] J. S. Choi, H. Lee, R. Elmasri, and D. W. Engels, "Localization systems using passive UHF RFID," in *Proceedings of the 5th International Joint Conference on International Conference on Networked Computing, International Conference on Advanced Information Management and Service, and International Conference on Digital Content, Multimedia Technology and its Applications (NCM '09)*, pp. 1727–1732, August 2009.

[6] S. Azzouzi, M. Cremer, U. Dettmar, R. Kronberger, and T. Knie, "New measurement results for the localization of UHF RFID transponders using an Angle of Arrival (AoA) approach," in *Proceedings of the 5th IEEE International Conference on RFID (RFID '11)*, pp. 91–97, April 2011.

[7] S. Azzouzi, M. Cremer, U. Dettmar, T. Knie, and R. Kronberger, "Improved AoA based localization of UHF RFID tags using spatial diversity," in *Proceedings of the 2nd IEEE RFID Technologies and Applications Conference (RFID-TA '11)*, pp. 174–180, September 2011.

[8] J. Hightower, G. Borriello, and R. Want, "SpotON: an indoor 3D location sensing technology based on RF signal strength," Tech. Rep., University of Washington, 2000.

[9] M. N. Lionel, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: indoor location sensing using active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.

[10] Y. Zhao, Y. Liu, and L. M. Ni, "VIRE: Active RFID-based localization using virtual reference elimination," in *Proceedings of the 36th International Conference on Parallel Processing in Xi'an (ICPP '07)*, September 2007.

[11] R. De Silva and P. Gonçalves, "Enhancing the efficiency of active RFID-based indoor location systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '09)*, April 2009.

[12] C. Wang, H. Wu, and N.-F. Tzeng, "RFID-based 3-D positioning schemes," in *Proceedings of the 26th IEEE International Conference on Computer Communications*, 2007.

[13] M. Bouet and G. Pujolle, "A range-free 3-D localization method for RFID tags based on virtual landmarks," in *Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, September 2008.

[14] J. S. Choi, H. Lee, D. W. Engels, and R. Elmasri, "Passive UHF RFID-based localization using detection of tag interference on smart shelf," *IEEE Transactions on Systems, Man and Cybernetics C: Applications and Reviews*, vol. 42, no. 2, pp. 268–275, 2012.

[15] W. Alsalih, K. Ali, and H. S. Hassanein, "A power control technique for anti-collision schemes in RFID systems," *Computer Networks*, vol. 57, no. 9, pp. 1991–2003, 2013.

[16] K. Ali, H. Hassanein, and A.-E. M. Taha, "RFID anti-collision protocol for dense passive tag environments," in *Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN '07)*, pp. 819–824, October 2007.

[17] W. Alsalih, K. Ali, and H. Hassanein, "Optimal distance-based clustering for tag anti-collision in RFID systems," in *Proceedings of the 33rd IEEE Conference on Local Computer Networks (LCN '08)*, pp. 266–273, October 2008.

[18] K. Chawla, G. Robins, and L. Zhang, "Object localization using RFID," in *Proceedings of the IEEE 5th International Symposium on Wireless Pervasive Computing (ISWPC '10)*, pp. 301–306, May 2010.

[19] R. Kronberger, T. Knie, R. Leonardi, U. Dettmar, M. Cremer, and S. Azzouzi, "UHF RFID localization system based on a phased array antenna," in *Proceedings of the IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting (APSURSI '11)*, pp. 525–528, July 2011.

[20] J. Zhou, H. Zhang, and L. Mo, "Two-dimension localization of passive RFID tags using AOA estimation," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC '11)*, pp. 511–515, May 2011.

[21] W. H. Hung, H. Lam, Y. W. Fung, D. C. Wei, and K.-L. Wu, "An active RFID indoor positioning system using analog phased array antennas," in *Proceedings of the Asia-Pacific Microwave Conference (APMC '11)*, pp. 179–182, December 2011.

[22] L. M. Ni, D. Zhang, and M. R. Souryal, "RFID-based localization and tracking technologies," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 45–51, 2011.

[23] 2013, Convergence Systems Limited, http://www.convergence.com.hk/.

[24] 2014, GAO RFID, http://www.gaorfid.com/.

[25] 2013, Datasheet, http://www.analog.com/static/imported-files/data_sheets/AD8302.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 927659, 13 pages
http://dx.doi.org/10.1155/2014/927659

Research Article

Energy Efficient and Load Balanced Routing for Wireless Multihop Network Applications

Vahid Nazari Talooki,[1] Jonathan Rodriguez,[1,2] and Hugo Marques[1]

[1] *Instituto de Telecomunicações (IT), Campus Universitário de Santiago, P-3810-193 Aveiro, Portugal*
[2] *Universidade de Aveiro, Aveiro, Portugal*

Correspondence should be addressed to Vahid Nazari Talooki; vahid@av.it.pt

Received 4 December 2013; Accepted 31 January 2014; Published 6 March 2014

Academic Editor: Christos Verikoukis

Wireless Sensor Networks (WSNs) can benefit from ad hoc networking technology characterized by multihop wireless connectivity and infrastructure less framework. These features make them suitable for next-generation networks to support several applications such M2M applications for smart cities and public safety scenarios. Pivotal design requirements for these scenarios are energy efficiency, since many of these devices will be battery powered placing a fundamental limit on network, and specifically node lifetime. Moreover, the way in which traffic is managed also influences network lifetime, since there is a high probability for some nodes to become overloaded by packet forwarding operations in order to support neighbour data exchange. These issues imply the need for energy efficient and load balanced routing approaches that can manage the network load and not only provide reduced energy consumption on the network but also prolong the network lifetime providing robust and continuous. This work proposes a new energy efficient and traffic balancing routing approach that can provide a weighted and flexible trade-off between energy consumption and load dispersion. Simulation results show that the proposed protocol achieves high energy efficiency, decreases the percentage of failed nodes due to lack of battery power, and extends the lifespan of the network.

1. Introduction

WSN is a key enabling technology for next-generation networks, having significant application in connecting a multitude of wireless tiny sensors or being able to operate in a more advanced mode by locally connecting different "things" of different capabilities, such as Personal Device Assistance (PDAs), Mobiles, and laptops, making then an ideal solution for next-generation networks. WSNs technology has a profound effect on our everyday life due to its inherent features such as being ubiquitous in nature and offering an inexpensive alternative to traditional networks.

WSNs are more specific use-cases for ad hoc wireless networks that today are autonomous in nature and support flexible topologies to deliver fast and everywhere connectivity. A self-configuring WSN, constituting distributed autonomous wireless nodes, can benefit from a multihopping ad hoc approach, in which nodes operate not only as a transceiver but also as a router and forward packets to other nodes in the network which may not be within direct transmission range of each other. Therefore, by operating in "ad hoc" mode, packet travelling from a source node toward a destination node may pass through multiple nodes to reach its destination.

A key enabler in ad hoc networks is a routing approach that needs to be robust and low in complexity to support end-to-end connectivity in highly dynamic operating environments. However, the effect of node/user mobility, dynamic topologies, frequent link breakages in the communication path, limitation on nodes resources such as battery power, and lack of central point such as base stations or servers means that routing in ad hoc networks can be a very challenging issue. Beside these aforementioned issues, there are also several qualities of service (QoS) metrics that should be considered in a communication, such as data throughput, delay, energy efficiency, traffic balancing, and the protocol overhead. Furthermore, due to the distributed and cooperated nature of ad hoc networks, attributes such as the number,

velocity, and the mobility pattern of mobile nodes may affect the QoS metrics provided by the routing protocol.

Providing a concrete routing solution that is energy efficient, but on the other hand able to deliver adequate QoS, is challenging, especially in a highly mobile environment. Usually each protocol tries to focus only on one or some of these QoS metrics. Different approaches reactive and proactive have been explored, and each has distinctive advantages.

Ad Hoc On-demand Distance Vector version 2 (AODVv2) protocol [1] is a well-known routing protocol presented by Mobile Ad Hoc NETwork (MANET) group of Internet Engineering Task Force (IETF) [2]; it applies a uniformed message format and has inherent features to support many required QoS metrics, but it is limited in terms of energy and traffic management. We exploit AODVv2 as a fundamental building block and go beyond this by designing and implementing energy and traffic aware capability. This work proposes an Energy Efficient Ad Hoc On-demand Distance Vector version 2 (E2AODVv2) that can have several applications specifically in WSNs. We implemented several simulation scenarios to test the capability of E2AODVv2, where results have shown that our proposed approach achieves higher performance in terms of energy consumption and load balancing compared to the baseline routing protocols.

This work is structured as follows: Section 2 presents the literature review and related works, Section 3 describes the current implementation of the AODVv2 protocol, Section 4 introduces the new proposed E2AODVv2 routing protocol, Section 5 presents the analysis and simulation results, and finally Section 6 presents conclusions and future works.

2. Related Works

2.1. Ad Hoc Routing Protocols Category. Many routing algorithms have been proposed for ad hoc networks in the literature. These routing protocols can be divided into several categories based on various criteria. In this section, we briefly review these categories and provide one sample from each category.

2.1.1. Flat Routing Protocols. Flat routing protocols in ad hoc networks adopt a flat addressing scheme which means all nodes participating in the routing process play an equal role. Flat routing protocols may generally be classified into two main categories.

(A) *Proactive Protocols.* This type of protocols attempts to find and maintain consistent, up-to-date routes between all source-destination pairs regardless of the use or need of such routes. Therefore, each node maintains one or more tables to store routing information (table driven protocols). Proactive protocols require periodic control messages to maintain routes up to date for each node. Routing techniques are either link-state or distance vector (or a mixture of both) [3, 4].

(i) Distance Vector (DV) based routing protocols: Destination Sequenced Distance Vector (DSDV) [5] is a proactive table-driven protocol based on the classical Bellman-Ford algorithm. All nodes try to find all paths to the possible destination nodes in a network and to save them in their routing tables.

(ii) Link-State Based (LS) routing protocols: Optimized Link-State Routing Protocol (OLSR) [6] uses selected nodes called multipoint relays (called MPRs) for routing operations in order to forward broadcast messages during the flooding process. Link-state information is generated only by MPRs and this information is then used for route calculation.

(B) *Reactive Protocols.* Routes are created only when a source node requests them (on-demand protocols). Forwarding process is accomplished according to two main techniques.

(i) Source routing protocols: Dynamic Source Routing (DSR) [7] is a source routing protocol and requires the sender to know the entire route to the destination. It is based on route discovery and route maintenance process. Discovered routes will be cached in the relative nodes.

(ii) Hop-by-hop routing protocols: like Ad Hoc On-demand Distance Vector (AODV) [8], it uses the on-demand mechanism of route discovery and route maintenance from DSR and also a mechanism for the hop-by-hop routing and sequence number. Per each destination, AODV creates a routing table, while DSR uses node cache to maintain routing information.

2.1.2. Hierarchical Routing Protocols. In hierarchical routing protocols, the network's nodes are divided into clusters. Each cluster has an admin entity (head cluster) which is responsible for routing inside that cluster; therefore, in this context, nodes have different responsibilities and importance in routing algorithm. Also, the cluster head election can be a dynamical and distributed operation. Zone Routing Protocol (ZRP) [9] is a hybrid/hierarchical routing protocol that takes advantage of both proactive and reactive schemes by creating overlapped zones based on the separation distances between wireless nodes. ZRP tries to limit the flooding area per each node by assigning a routing zone to that node [10].

2.1.3. Geographical Position Based. These types of protocols assume that each node in the network is aware of its own location and the status of its one-hop neighbors, for example, via a Global Positioning System (GPS). In the Greedy Perimeter Stateless Routing (GPSR) [11] protocol, in any communication between a source-destination pair of nodes, the source node is aware of the destination node's location. All one-hop neighbors exchange beacon control messages between each other to simultaneously update their routing tables and limit control message overhead [10].

2.1.4. Hybrid Protocols. Hybrid protocols, by mixing different routing algorithms, try to take the advantage of previously mentioned protocols to reach to the highest efficiency. Temporally-Ordered Routing Algorithm (TORA) [12] is a highly adaptive loop-free distributed routing algorithm based

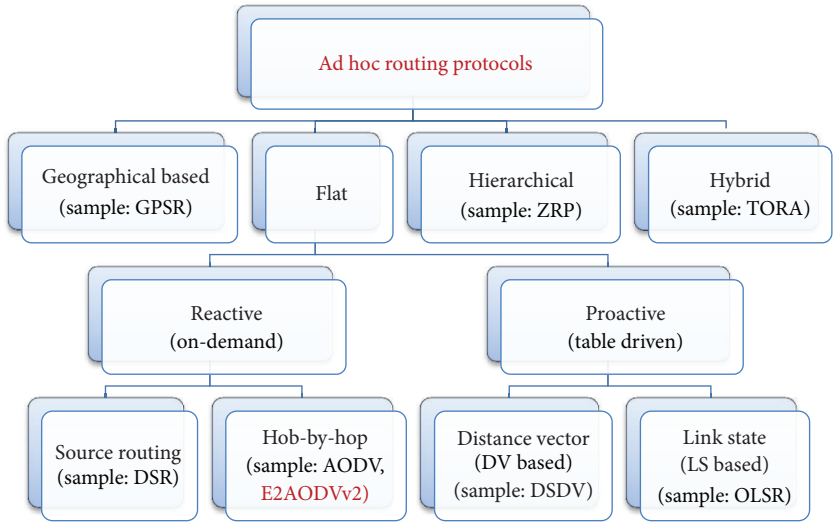


Figure 1: A structure of ad hoc routing protocols categories.

on the concept of link reversal and it applies a reactive (on demand) source routing scheme. The key design concept of TORA is the reduction of control messages to a very small set of topological changes by performing three basic functions: route creation, route maintenance, and route deletion. All nodes in TORA use a "height" metric to establish a direct acyclic graph (DAG) rooted at the destination [13].

A category of current ad hoc routing protocols is shown in Figure 1. Also, a survey on current routing algorithms for wireless ad hoc networks can be found in [10]. The new proposed routing protocol, the so-called "E2AODVv2," can be conveniently positioned as shown by Figure 1.

2.2. Power-Aware Routing Protocols. Hard QoS routing guarantee in ad hoc networks is a Nondeterministic Polynomial time- (NP-) Complete problem [14]. Therefore, the current works focus on providing a soft QoS routing guarantee for ad hoc networks as a more realistic solution. In this paper, we consider battery power consumption and traffic balancing as a measurement of efficiency.

Most of current routing protocols, proposed in MANET group of IETF [2], consider the path length metric when choosing the best route between a source (S) and a destination node (D). However, this approach may, in most cases, minimize the end-to-end delay in a communication between the source and the destination node, but it may not be adept to handle other QoS metrics such as energy efficiency and load balancing, because they do not consider the node residual energy as a criterion in the route selection process [15].

The dynamic topology in ad hoc networks implies that some nodes may relay more traffic than others, mainly because of their location in the network; these hot spots will consume their energy reserves sooner than the others. Unbalanced battery power consumption in the network nodes can leads to early node failure rate a, network partitioning, and to a reduction in the route reliability. Traffic concentration on these nodes may increase radio jamming, delay, and packet loss. Also, since the majority of the network traffic can potentially pass these nodes, they can become an important target for attackers.

Power-aware routing algorithms aim to deliver new routing paths that take into account energy as a metric [16]. Gomez et al. [17] add new intermediate nodes to a route from a source to a destination to reduce the overall required transmission power of the intermediate nodes. Lindgren and Schelen [18] improve the AODV routing protocol in terms of energy efficiency, by selecting paths through Power Base Stations (PBSs) instead of through normal nodes. Edwards et al. [19, 20] present two power-aware extensions of the original DSR ad hoc routing protocol that use energy aware metrics such as the remaining battery power to decide which nodes should participate more often in packet forwarding. Load balancing and homogeneous distribution of battery power consumptions between networks nodes are other solutions to achieve a power-aware ad hoc routing algorithm [21].

However, most of current power-aware approaches lead to some common drawbacks, such as increased delay and increase in the number of required control messages that should be created by the routing protocol to deliver users' data packets (called Normalized Routing Load), limited scalability, among others.

3. Revised Ad Hoc On-Demand Distance Vector (AODVv2)

AODVv2 is a successor to the AODV that is being developed by IETF MANET; prior to the 26th revision [1], AODVv2 was called DYnamic MANET On-demand (DYMO).

3.1. Protocol Description. AODVv2 tries to simplify the current reactive protocols, such as DSR and AODV, and simultaneously conserves their two main well-known routing operations: route discovery and route maintenance [22]. It has multipath capability (optional) and is also a hop-by-hop routing protocol. Therefore, intermediate nodes, located in a route between a source and a destination node, are able

to extract additional information from the traversing control packets. AODVv2 applies a standard generalized MANET Packet/Message format [23] for control packets: a uniformed message format, called Routing Message (RM), is used for all control messages and routing packets. Another interesting feature of AODVv2 is the capability to Internet Protocol (IP) [24] versions 4 and 6 (IPv4 and IPv6, resp.) which can be considered an essential feature for next-generation networks.

Several works analysed the performance and benefits of AODVv2 [22, 25, 26]. To improve the performance of AODVv2 [22, 25, 26], Kretschmer et al. [27] focused on reducing the delay and on increasing the packet delivery ratio, by finding routes with high probability to be used in future. Also, there are few works, such as [25], that try to extend a multipath version of AODVv2 to switch between different paths from source to destination in order to balance the battery power consumption and the traffic of the nodes on different routes. However, these works do not provide a mechanism for handling both load balancing and energy efficiency in highly dynamic topologies.

3.2. Routing Operation. All routing operation in AODVv2 can be categorized into three phases.

3.2.1. Route Request Phase. In a communication between a source node (S) and a destination node (D), S originates a Route Message (RM), called ROUTE REQUEST (RREQ) message, and broadcasts this to all of its neighbours. These RREQs are then flooded in the network until they reach D. An intermediate node which does not know the route to D should forward the RREQ to its neighbours. The intermediate node will also drop the repeated request for the same destination and will not forward them anymore, as shown by Figure 2. RM has several fields such as current node address, next node address, HopLimit (the default value is 10 based on [1]), Target (destination node address if it is an RREQ message), and the Origin (source node address if it is an RREQ message) [1]. By tracing the RREQ traversed path (called *accumulated path*), each node, such as intermediate or destination node, which receives an RREQ message, can extract routing information. Figure 2 shows the RREQ phase in AODVv2 routing protocol for a sample network.

3.2.2. Route Reply Phase. As shown by Figure 3, when RREQs finally reach the destination node, another RM, which is called ROUTE REPLY (RREP), will be originated by the destination node. The intermediate nodes which have, in their routing table, an entire route towards that destination node also can immediately reply to the RREQ originated by the source node (known as *gratitude reply* and it is an optional feature of AODVv2). Both the RREQ and the RREP have the same uniform structure.

3.2.3. Routing Operation: Route Error Phase. An intermediate node may originate an RM, called an ROUTE ERROR (RERR), in the two main scenarios. In the first case, the intermediate node does not have a valid route for the destination of a received data packet and consequently the packet is

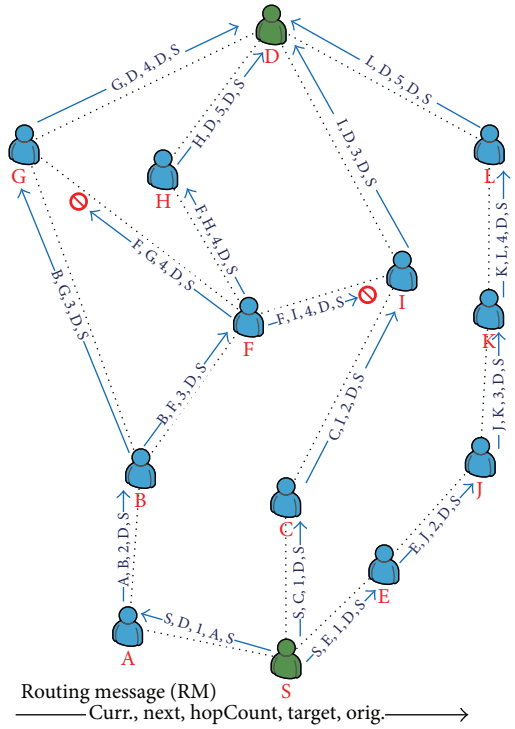


Figure 2: A sample of RREQ phase in AODVv2.

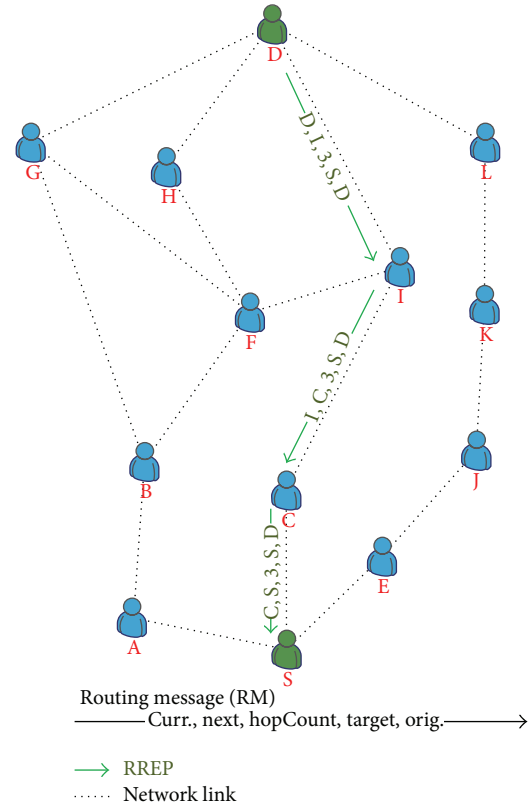


Figure 3: A sample of RREP phase in AODVv2.

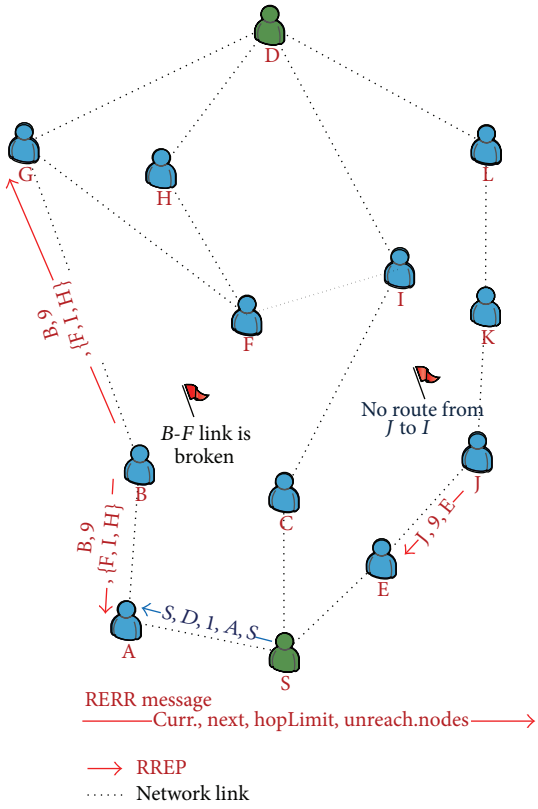


Figure 4: A sample of RERR phase in AODVv2.

undeliverable. In the second case, the intermediate node detects a link breakage, as shown in Figure 4.

4. Energy Efficient AODVv2 (E2AODVv2)

The inherent benefits of AODVv2 due to standardized control packets following IETF uniform packet formats and the capability to support IPv6 suggest that this protocol will have a strong legacy in ad hoc networks. Therefore, if we are to pursue energy efficient operation for routing in ad hoc networks, then exploring AODVv2 as the fundamental building block can be potentially a springboard for promoting significant energy savings in the network. In this section, we describe our proposed approach to increase the energy efficiency of AODVv2.

4.1. Routing Messages in E2AOVv2. E2AODVv2 introduces two new fields for each routing message which are discussed in this section.

4.1.1. Energy Field. Suppose that, for a communication between a source and a destination node, there are N routes and the number of nodes in the ith route, called Route$_i$, is M_i (as shown in Figure 5). In E2AODVV2, each kth node in Route$_i$, has a battery power (BP$_k$) level quantified as 16 different values (from 0 to 15).

Moreover, there is a Critical Battery Power Level (CBPL) whose default value is 3 (CBPL = 3). Therefore node k in

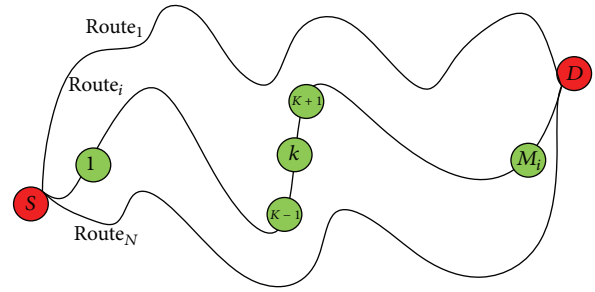


Figure 5: Multiple routes between source and destination.

Table 1: Battery power level and traffic parameter of all nodes of the network in Figure 6.

Node	Battery power (BP)	Traffic parameter (TP)
A	6	0.7
B	8	0.5
C	2	0.8
F	12	0.3
I	10	0.6
G	5	0.9
H	14	0.4

Route$_i$, which has a BP of less than CBPL, is suffering from energy depletion and is considered as a critical node in terms of battery power (i.e., BP$_k$ < CBPL). Critical nodes should not be selected for packet forwarding in E2AODVv2.

The *Energy* field of RREQ in E2AODVv2 has three cells: totbat, MinBat, and CritBat. TotBat$_i$ is the summation of the total battery power level of all nodes of Route$_i$, as shown by (1). The MinBat$_i$ cell in the energy field of RREQ$_i$ shows the minimum value of BP for all nodes in Route$_i$, whilst CritBat$_i$ shows the number of nodes which have a BP less than CBPL in Route$_i$:

$$\text{TotBat}_i = \sum_{k=1}^{k=M_i} \text{BP}_k. \quad (1)$$

Figure 6 shows a network based on the previous sample that has a simpler topology. In this example, there are 8 nodes. The battery power level and the traffic parameter of all these nodes are given in Table 1 (the traffic parameter will be introduced later in this paper). Based on Table 1, the energy and the traffic parameters of all paths of the network in Figure 6 are calculated and given by Table 2.

For example, as the first column of Table 2 shows, the first path (i = 1) from S to D (i.e., S-C-I-D) has one node with a critical battery level (CritBat = 1); the minimum battery power level of all nodes in this path is 2 which belongs to node C, as shown in Table 1 (MinBat = 2); the total battery power level of all nodes in this path is 12 (TotBat = 12), and the number of intermediate nodes in this path is 2 (M = 2).

For this network, the original AODVv2 chooses the shortest path (i.e., the first path S-C-I-D); however, the new

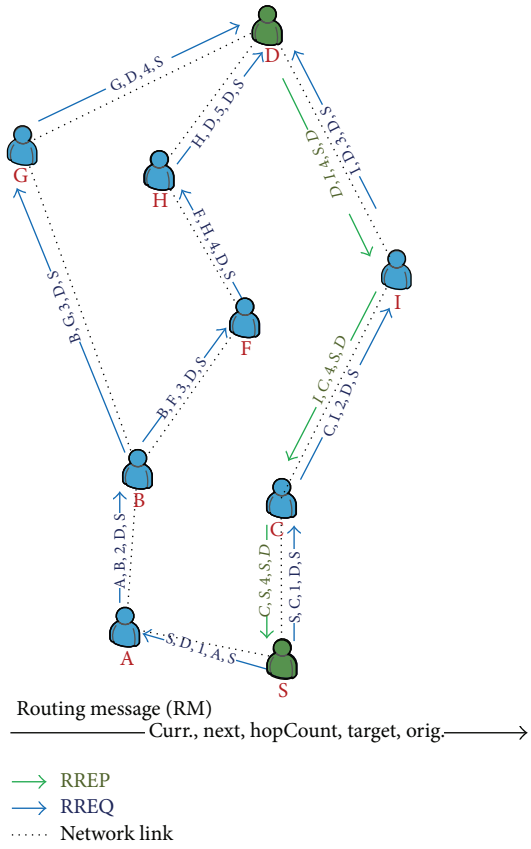


Figure 6: D originates route replies (RREP) toward S in AODVv2.

Table 2: The energy and traffic features of all these nodes of the network in Figure 6.

Path	S-C-I-D	S-A-B-F-H-D	S-A-B-G-D
Path number	$i = 1$	$i = 2$	$i = 3$
CritBat	1	0	0
TotBat	12	40	19
MinBat	2	6	5
M	2	4	3
TotTra	1.4	1.9	1.8
MaxTra	0.8	0.7	0.7

proposed protocol, E2AODVv2, will choose the second path (i.e., S-A-B-F-H-D), which is a stronger alternative in terms of energy efficiency, and even traffic load which is discussed later on.

4.1.2. Traffic Field. In Figure 5, if the queue size of the interface of node k located in Route$_i$ is AQ$_k$ and the maximum queue size is MQ$_k$, then the traffic parameter of node k in E2AODVv2 (i.e., TP$_k$) will be calculated based on

$$\mathrm{TP}_k = \frac{\mathrm{AQ}_k}{\mathrm{MQ}_k}. \tag{2}$$

The Traffic field of RREQ has two cells: TotTra and MaxTra. The first one is the summation of all traffic parameters in Route$_i$, as defined by

$$\mathrm{TotTra}_i = \sum_{k=1}^{k=M_i} \mathrm{TP}_k. \tag{3}$$

The second one, that is, MaxTra$_i$, is the maximum TP values of Route$_i$. The traffic parameters of all nodes of the network, which is presented in Figure 6, are given by Table 1, whilst the traffic parameters of all paths are given in Table 2.

For example, as the first column of Table 2 shows, for the first path ($i = 1$) from S to D (i.e., S-C-I-D), the total traffic parameter of all intermediate nodes is 1.4 (i.e., TotTra = 1.4) and the maximum traffic parameter is 0.4 (i.e., MaxTra = 0.4) which belongs to node C as shown in Table 1. Also, M is the number of intermediate (relay) nodes in a path; for example, the first path S-C-I-D has two intermediate nodes; therefore $M = 2$.

Therefore, in E2AODVv2, when the intermediate node k receives RREQ, it updates TotTra$_i$ and MaxTraiof the traffic field of the RREQ and forwards the RREQ toward the next node. A sample node routing table is shown in Table 3.

4.2. Route Selection Process. In E2AODVv2, when a destination node receives several RREQs from different routes, as in Figure 5, it runs a route selection process to determine the best route in terms of *energy* and *traffic* parameters.

4.2.1. Calculating Energy Parameter of a Route in E2AODVv2. The energy parameter of Route$_i$, called $E(i)$, indicatesthe priority of the Route$_i$, in terms of the battery power level as presented by

$$E_i = \frac{\mathrm{TotBat}\,(i)}{M_i \times \mathrm{InitialBat}}, \tag{4}$$

where M_i is the number of nodes in Route$_i$, TotBat(i) is the total battery power level of all nodes in Route$_i$, as presented in (1), and InitialBat is the initial battery power level of a node (which is preset to the same value for all nodes). In some scenarios, a route may have a few nodes with a very low energy level, but a high overall energy level; this route should be avoided due to these bottleneck nodes. Therefore, the negative impact of MinBat(i) (which is the minimum battery power level of Route$_i$) should be applied to $E(i)$, as given by

$$E_i = \frac{\mathrm{TotBat}\,(i) \times \mathrm{MinBat}\,(i)}{M_i \times \mathrm{InitialBat}^2}. \tag{5}$$

As mentioned before, CritBat(i) identifies the number of nodes in Route$_i$ which have reached to a critical battery level and hence should be avoided from packet relaying functions (otherwise, the route is likely to break down). A large value for CritBat(i) triggers an alert that this route has several nodes with a critical battery level. Consequently, $E(i)$ can be revised to reflect also the negative impact of these nodes in Route$_i$. The final $E(i)$ of a route in E2AODVv2 will be calculated based

Table 3: A sample of routing table of a node in E2AODV2.

Dest. add.	Route seq. num.	Hop count	Next hop	Lifetime (ms)	Last used (ms)	Forward flag	Broken flag	⋯
D	112	3	I	30000	10000	T	F	⋯
E	36	2	S	20000	15000	T	F	⋯
K	76	4	I	50000	1000	T	F	⋯
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮

on (6). When a destination node receives an RREQ, it can calculate E_i, via the energy field of the RREQ:

$$E_i = \frac{\text{TotBat}(i) \times \text{MinBat}(i)}{M_i \times \text{InitialBat}^2 \times (\text{CritBat}(i) + 1)}. \quad (6)$$

4.2.2. Calculating Route Traffic Parameter in E2AODVv2. A route with a lower traffic metric cost has a higher priority in the routing process of E2AODVv2. Nevertheless, a route may have a low overall traffic metric even if one of its nodes is overloaded with traffic, thus creating a bottleneck, and should be potentially avoided. The traffic parameter of Route$_i$ is shown by $T(i)$ and is calculated by

$$T_i = \frac{\text{TotTra}(i) \times \text{MaxTra}(i)}{M_i}, \quad (7)$$

where M_i is the number of nodes in Route$_i$. When a destination node receives an RREQ, it can calculate T_i via the traffic field of the RREQ.

4.2.3. Precedence Function. The function that determines the precedence of Route$_i$ is given by RoutePrio(i). This function depends on the energy and the traffic parameters of Route$_i$ that are calculated by (6) and (7), respectively:

$$\text{RoutePrio}(i) = \frac{E_i}{T_i}. \quad (8)$$

However, we can choose a trade-off between energy efficiency and traffic balancing, but this function is flexible and can be customized by giving a higher weight to the energy or traffic parameter.

4.2.4. Routing Behaviour of Nodes in E2AODVv2. In E2AODVv2, each node shows different routing behaviour that depends on the role of the node in the communication.

(i) Source node: when a source node wants to send an RREQ, it initializes the values of the energy and the traffic fields (they will be set to zero).

(ii) Intermediate nodes: these update the energy and traffic fields accordingly: (i) each node who receives an RREQ adds its own BP to TotBat value; (ii) if the current node k in Route$_i$, has a BP$_k$ less than MinBat, then MinBat will be updated by BP$_k$ (i.e., if BP$_k$ < MinBat$_i$ then MinBat$_i$ = BP$_k$); (iii) if BP$_k$ < CBPL, then CritBat will be increased by one- (iv) each node, which receives an RREQ and adds its TP$_k$ number to the TotTra number; (v) if TP$_k$ > MaxTra, then MaxTra will be updated by TP$_k$. However, the intermediate node in traditional AODVv2 will drop the repeated request for the same destination and will not forward them anymore, as Figure 2 shows, but, in our protocol, the intermediate nodes will forward these RREQs as well, as shown in Figure 7.

(iii) Destination nodes: when a destination node receives n RREQs from different n routes, by extracting the required information from RREQ, it can calculate the Route Priority for all these RREQs and finally the route which has the best Route Priority will be chosen.

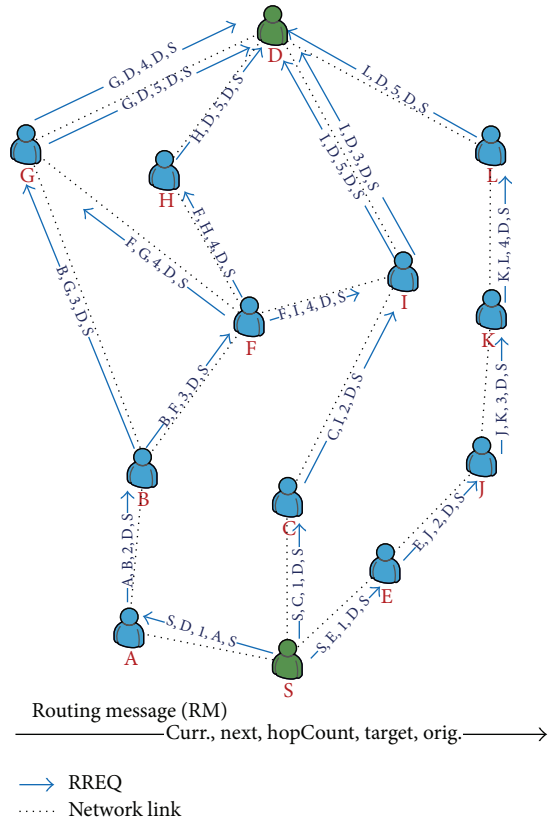


Figure 7: A sample of RREQ phase in E2AODVv2.

Figure 8 shows the role of the source, the intermediate, and the destination nodes in a communication. Each route request, such as RREQ$_i$, should be initialized by a source node. It will pass from one node to another within the route, and each node has the opportunity to update the energy and traffic fields accordingly (where each route will have M_i

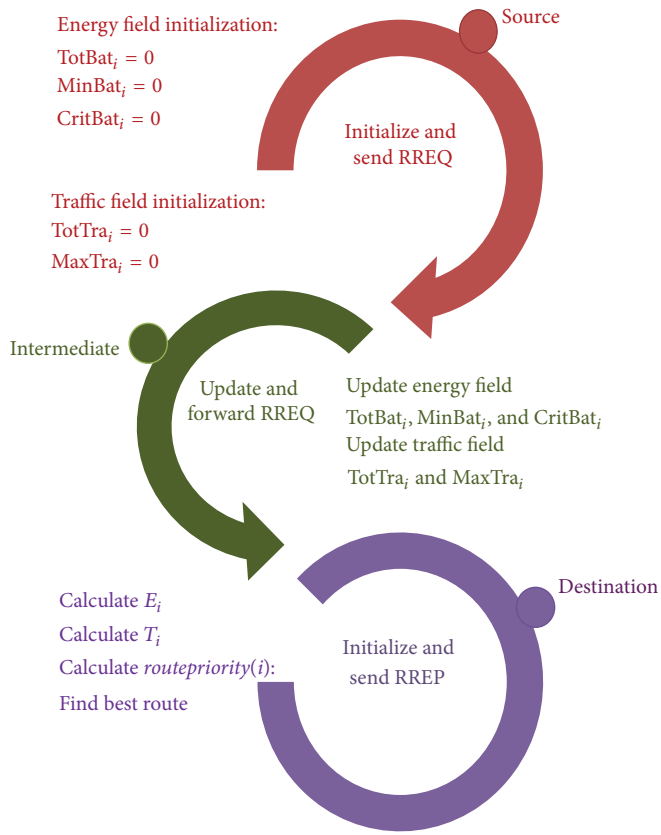


Figure 8: RREQ and RREP processes: the routing behaviour of source, intermediate, and destination nodes in E2AODVv2.

intermediate nodes). It finally reaches the destination node which is responsible for running the Precedence Function, finding the best route and initializing the RREP. Therefore, via this distributed mechanism, the nodes in E2AODVv2 can cooperatively balance the load in the network, in terms of traffic and battery power consumption.

5. Analysis of E2AODVv2

5.1. Analytical Analysis. As we discussed in Section 4.1, the traditional AODVv2 for the network in Figure 6 chooses the first path ($i = 1$) from S to D (i.e., S-C-I-D). The battery power level and the traffic parameter of all these nodes are given in Table 1. Also, based on Table 1, the energy parameter, the traffic parameters, and the priority of all paths of the network in Figure 6 are calculated and given in Table 4. Our proposed E2AODVv2 protocol, based on Table 4, chooses the second path (i.e., S-A-B-F-H-D), which provides equal priority towards reducing energy and balancing traffic load. Therefore, E2AODVv2 send RREP messages through this optimized path as shown by Figure 9.

5.2. Evaluating Proposed Protocol. We use Two-Ray Ground Reflection Model which considers both the direct path and a ground reflection path between two mobile nodes. Also all nodes in the network move based on the Random Way Point Mobility Model. In this movement model, a mobile node starts its travel from a random location inside the topology area after pausing for a certain period of time (called "*pause time*"). After the initial pause time, the mobile node chooses another random location inside the topology area and moves toward this new location by a speed that is uniformly distributed between a predefined minimum and maximum speed. Upon arrival, the mobile node pauses again for the *pause time* and repeats the previous process again till the simulation time is expired [3, 28].

The simulation results presented in this paper were obtained using the *ns-2* simulator [29]. Traffic sources are

Table 4: The energy and traffic features of all paths in the network of Figure 9.

Path	S-C-I-D	S-A-B-F-H-D	S-A-B-G-D
Path number	$i = 1$	$i = 2$	$i = 3$
CritBat	1	0	0
TotBat	12	40	19
MinBat	2	6	5
M	2	4	3
E	0.026	0.26	0.14
TotTra	1.4	1.9	1.8
MaxTra	0.8	0.7	0.7
T	0.56	0.3325	0.63
RoutePrio (E/T)	0.046	0.781	0.22
Best path	×	√	×

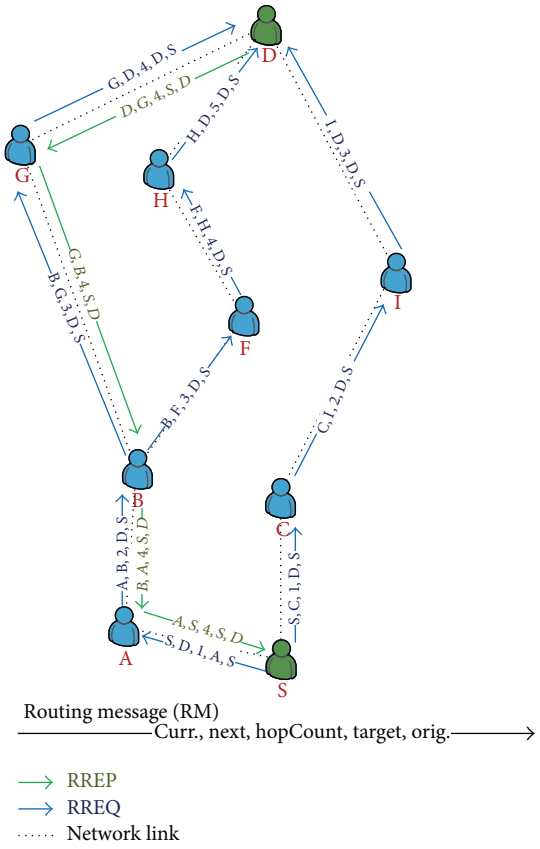


Figure 9: A sample of RREP phase in E2AODVv2.

CBR (Constant Bit Rate) and the packet sending rate at the source nodes is 8 packets per second. Table 5 presents a summary of the parameters for the simulated scenarios (movement and traffic files).

5.2.1. Performance Metrics. We evaluated the proposed protocol using five metrics: (1) balancing energy consumption, (2) scalability, (3) network lifetime, (4) node's failure level, and (5) jitter.

(*I*) *Balancing Energy Consumption*. This metric will allow measuring the effectiveness of the energy balancing algorithm used by E2AODVv2. Let us denote the energy load of all the nodes in the network consisting of N nodes; is a set $\mathrm{EL} = \{el(i) \in \mathrm{El} : i = 1, \ldots, N\}$,

where element of this set, for example, el(i), is calculated as ration of the consumed energy node i to the total aggregate energy consumed in all the nodes, including node i. That is,

$$\mathrm{el}(i) = \frac{\text{Consumed Energy}(i)}{\text{Total Consumed Energy}}. \tag{9}$$

Having calculated all N elements of the EL, we can now calculate the standard deviation of EL for a network consisting of N nodes (i.e., σ_{EL}^N). The value of σ_{EL}^N will be our energy balancing metric for comparing different protocols; the smaller the σ_{EL}^N, the more effective the energy balancing capability.

Table 5: Parameters of movement models and communication model.

Parameters of movement model I, characterized by number of nodes	
Topology area	500 m × 500 m
Maximum mobility of nodes	5 m/s
Number of nodes	1, . . . , 20
Simulation time	300 s
Parameters of movement model II, characterized by simulation time	
Topology area	500 m × 500 m
Maximum mobility of nodes	5 m/s
Number of nodes	20
Simulation time	200, . . . , 1800 s
Parameters of traffic model	
Traffic sources	CBR
Data packets size	512 bytes
Sending rate	8 packets/second

(*II*) *Scalability*. Another interesting metric is the scalability of the proposed protocol in terms of load balancing, that is, if the proposed protocol can balance the load of the network when we increase the number of network nodes (i.e., N in our settings). For this purpose we change the number of nodes N (the network size) to a maximum number (which is 20 in our setting, based on Table 5) and calculate the related σ_{EL}^N for each network size.

(*III*) *Network Lifetime*. Network lifetime is another important metric that reflect the load balancing capability of the routing protocol; the bigger the lifetime, the more effective the energy balancing capability. To compare the lifetime of the proposed protocol with other protocols, we take two parameters into consideration: (1) the first failure time $\mathrm{FT}_{\mathrm{first}}$ (the time that the first node in the network ran out of battery power) and (2) last failure time $\mathrm{FT}_{\mathrm{last}}$.

(*IV*) *Node Failure Level*. Finally the last performance metric that we use is the node failure level, which determines the capability of E2AODVv2 in keeping nodes alive for longer durations. The node failure level, for a time window T, is the percentage of nodes in the topology that have failed due to a depleted battery. This value is calculated as follows:

$$\text{Node Failure Level} = \frac{\#\text{failed\_nodes\_in\_}T}{\#\text{nodes\_in\_topology}}. \tag{10}$$

(*V*) *Jitter*. In a data transmission between a pair of source and sink nodes, jitter is the variation in the time between packets arriving at the source node. Let us assume that, at time S_i, the source node sends packet P_i and the sink node receives it at time R_i. The jitter of packet P_i is calculated as follows:

$$\begin{aligned} \text{Jitter } P_i &= \left|(R_{i+1} - R_i) - (S_{i+1} - S_i)\right| \\ &= \left|(R_{i+1} - S_{i+1}) - (R_i - S_i)\right|. \end{aligned} \tag{11}$$

During the entire simulation time, in a communication between a pair of source and sink nodes, there are M streams

Table 6: Balancing energy consumption metric σ_{EL}^{N}, the first failure time $\mathrm{FT}_{\mathrm{first}}$ (s), and the last failure time $\mathrm{FT}_{\mathrm{last}}$ (s), for both distributed and concentrated traffic modes.

	$\sigma_{\mathrm{EL}}^{12}$ Distributed	$\sigma_{\mathrm{EL}}^{12}$ Concentrated	$\mathrm{FT}_{\mathrm{first}}$ Distributed (second)	$\mathrm{FT}_{\mathrm{first}}$ Concentrated (second)	$\mathrm{FT}_{\mathrm{last}}$ Distributed (second)	$\mathrm{FT}_{\mathrm{last}}$ Concentrated (second)
E2AODVv2	0.029	0.085	500	500	1700	1800
DSR	0.045	0.182	400	400	1700	1500
AODV	0.05	0.202	300	300	1400	1300

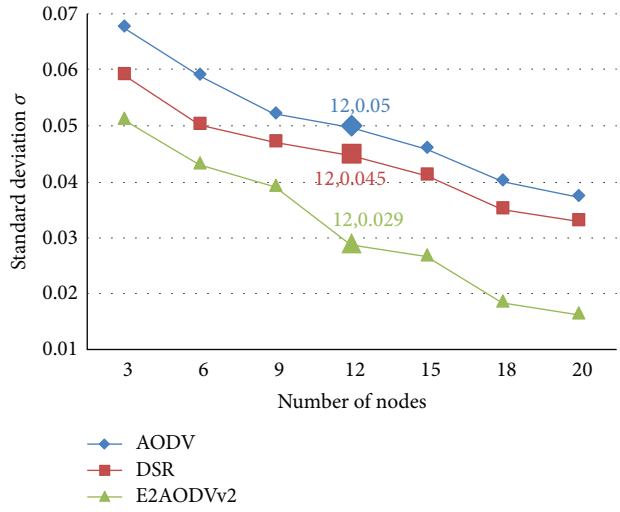


Figure 10: Balancing of energy consumption (σ_{EL}^{N}) and the scalability of protocols versus number of nodes (when distributing traffic amongst nodes).

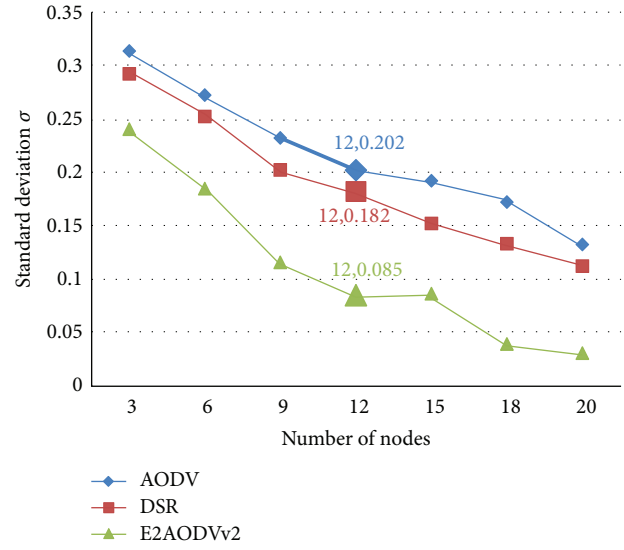


Figure 11: Balancing of energy consumption (σ_{EL}^{N}) and the scalability of protocols versus number of nodes (when traffic is concentrated in few nodes).

of packets, with each stream consisting of P packets. We studied the average jitter of all streams of data in the network.

5.2.2. Simulation Results. This section presents the simulation results for the comparison between two reactive protocols AODV and DSR that are considered baselines in this work, against our proposed E2AODVv2 protocol.

(*A*) *Balancing Energy Consumption and Scalability*. For this metric, simulation results are presented in terms of the capability of our approach to balance battery power (energy) consumption. The traffic model is the one presented in Table 5 with the following variants.

(i) Distributed traffic amongst nodes: traffic sources and destinations will be chosen randomly in time. This variant is shown by "Distributed" in Table 6.

(ii) Concentrated traffic with overloaded nodes: traffic destinations will be predefined (all traffic in the network goes toward those sink nodes). This variant is shown by "Concentrated" in Table 6.

The metric for energy consumption balancing in a network consisting of N nodes is σ_{EL}^{N} as discussed in Section 5.2.1. As a sample, the value of σ_{EL}^{N} when network consists of 12 nodes (i.e., σ_{EL}^{12}) is marked in Figure 10 per different protocols for the *distributed traffic* mode and similarly in Figure 11 for the *Concentrated traffic* mode. Also, Table 6 summarizes these sample values for σ_{EL}^{12}.

Furthermore, the value of σ_{EL}^{N} decreases as the network size N increases, demonstrating the scalability of the protocols as shown in Figures 10 and 11 for *Distributed* and *Concentrated* traffic modes, respectively.

For the first variant, *distributed traffic amongst nodes*, all protocols show a similar performance (Figure 10): increasing the number of nodes leads to improved energy balancing (lower standard deviation). This can be explained due to the chosen traffic model since multiple traffic source/destinations are chosen randomly throughout the simulation, and thus more nodes will participate in the routing process.

For the second variant, that targets concentrated *traffic with overloaded nodes*, we chose n nodes, amongst all network nodes (N), as destination nodes (sink nodes) for the generated traffic in the network. The number of n is set to 20% of N, which due to the many-to-one traffic pattern, creates overload conditions at destination nodes. Simulation results show that, in such scenarios, the standard deviation behaviour is similar to the previous variant but with the worst performance (Figure 11). However, as in the previous scenario, all three protocols improve in terms of performance as the number

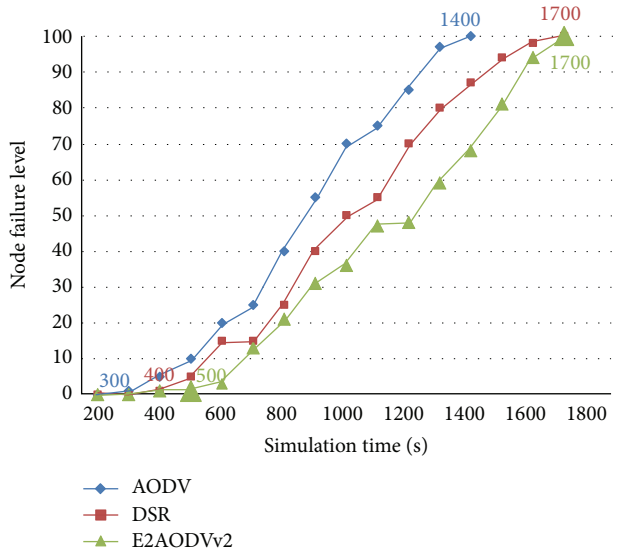


Figure 12: Percentage of failed nodes versus simulation time (when distributing traffic amongst nodes).

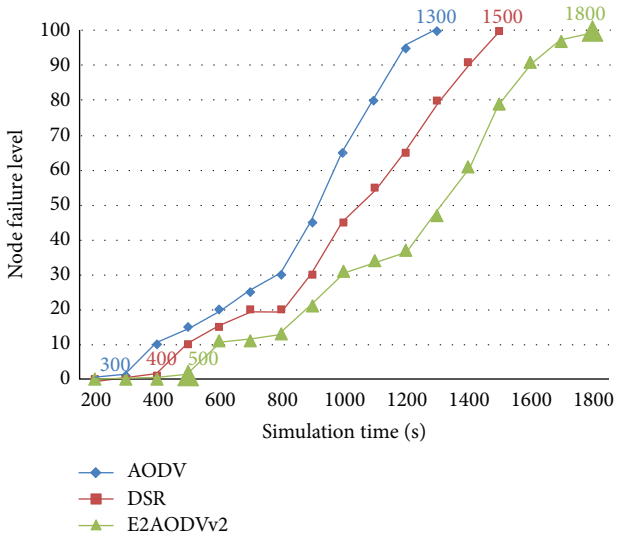


Figure 13: Percentage of failed nodes versus simulation time (when traffic is concentrated in a few nodes).

of nodes in topology increases (Figure 11). As both Figures 10 and 11 show, the proposed protocol E2AODVv2 reaches a higher performance in terms of battery power efficiency, balancing, and scalability in contrast to the baselines.

(*B*) *Network Lifetime and Node Failure Level*. This metric is measured using *movement model II* described in Table 5. Similar to the previous case, here again there are two different traffic models: *Distributed traffic amongst nodes* (Figure 12) and *Concentrated traffic with some overloaded nodes* (Figure 13). As discussed in Section 5.2.1, we use the first failure time FT_{first} and the last failure time FT_{last} as the network lifetime metrics. The values of FT_{first} and FT_{last} are marked in Figures 12 and 13 and also summarized in Table 6.

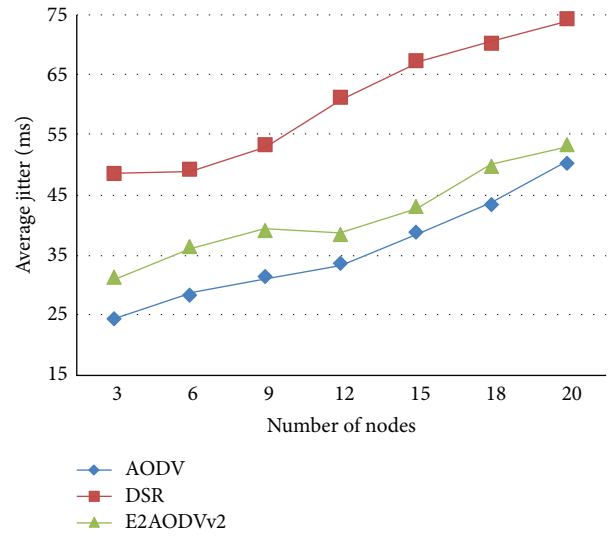


Figure 14: Jitter versus number of nodes.

As these results show, E2AODVv2 has always better value for both FT_{first} and FT_{last}. Also, as expected, the number of failing nodes increases with simulation time. When nodes use E2AODVv2, their lifetime gets extended due to the energy balancing capabilities of E2AODVv2 and shows a lower *Node failure level* (Figure 12). This feature is even more pronounced when traffic is concentrated in a few nodes (Figure 13).

(*C*) *Jitter*. Jitter has several sources such as network congestion, route changes, and delaying packets in the buffer queues of the intermediate nodes in a source-sink communication session. Usually a jitter buffer is used at sink nodes to counteract the jitter.

Jitter is a measurement for the quality of the communication; a small jitter indicates a high quality communication by a low latency. As Figure 14 shows, the performance of the proposed protocol is better than DSR and very close to the original AODV. This can be explained due to the more complicated algorithm of E2AODVv2 for choosing a route which leads to a higher delay.

Future work will extend the proposed approach in this paper for energy efficiency and load balancing to include other QoS metrics such as delay and data throughput by varying the topology area and velocity of nodes. Another metric that could be considered in the future works to obtain the required power for transmitting the data between two nodes is the signal to noise ratio (SNR). A high SNR value for a link between two neighbor nodes shows that we should find an alternative energy efficient link and eventually a path for the data transmission between a pair of source and sink nodes. This approach may also lead to the need for a cross-layer design [30] for our protocol.

6. Conclusion

Next-generation wireless sensor networks, constituting distributed autonomous wireless sensors and nodes, can benefit

from ad hoc networking technology. However a key requirement for this technology that should be satisfied is an energy efficient and load balanced routing protocol. This paper proposes a new energy efficient and traffic balancing routing protocol based on the well-known IETF AODVv2 protocol, a popular approach for routing in ad hoc networks. The protocol uses a standard generalized MANET Packet/Message format. The E2AODVv2 has been enhanced with battery power efficiency and balancing capability that can detect the nodes that reach a critical battery level in the network and switch the route in order to avoid network fragmentation and to achieve a higher network lifetime. The same behaviour is also fulfilled when bottlenecks are detected in a specific route in terms of traffic load. These additional functionalities are achieved without the need to create new disruptive approach in the protocol messages. E2AODVv2 achieves a higher performance with respect to energy consumption balancing, scalability, network lifetime, and the percentage of failed nodes in comparison to well-known baseline protocols such AODV and DSR and a jitter value very close to the AODV. As shown by the simulation results, the E2AODVv2 routing protocol specifically outperforms in scenarios where the load of the network is not balanced. Moreover, the proposed approach can be said to be technology agnostic in that it can be applied to many other reactive ad hoc protocols.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The research leading to these results has received funding from FEDER through the Fundação para a Ciência e Tecnologia [national reference ARTEMIS/0005/2012 - ACCUS] and the ARTEMIS JU [ART Call 2012 - 333020 ACCUS].

References

[1] IETF, "Dynamic MANET on-demand (AODVv2) routing, draft-ietf-manet-dymo-26," 2013, http://tools.ietf.org/html/draft-ietf-manet-dymo-26.

[2] MANET, "Working group in IETF," 2013, http://datatracker.ietf.org/wg/manet/.

[3] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, Dallas, Tex, USA, 1998.

[4] M. E. Illyas, "Security in wireless ad hoc networks," in *The Handbook of Ad Hoc Wireless Networks*, chapter 30, CRC press, Boca Raton, Fla, USA, 2003.

[5] C. E. Perkins and P. Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 4, pp. 234–244, 1994.

[6] A. Laouiti, P. Jacquet, P. Minet, L. Viennot, T. Clausen, and C. Adjih, "Optimized link state routing protocol (OLSR)," 2003.

[7] D. B. Johnson, D. A. Maltz, and J. Broch, "DSR: the dynamic source routing protocol for multihop wireless ad hoc networks," in *Ad Hoc NetworkIng*, pp. 139–172, Addison-Wesley, Boston, Mass, USA, 2001.

[8] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA '99)*, pp. 90–100, New Orleans, La, USA, February 1999.

[9] Z. J. Haas and M. Pearlman, "The Zone Routing Protocol (ZRP) for ad-hoc networks, IETF internet draft," July 2002.

[10] E. Alotaibi and B. Mukherjee, "A survey on routing algorithms for wireless ad-hoc and mesh networks," *Computer Networks*, vol. 56, no. 2, pp. 940–965, 2012.

[11] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, pp. 243–254, Boston, Mass, USA, August 2000.

[12] V. D. Park and M. S. Corson, "Highly adaptive distributed routing algorithm for mobile wireless networks," in *Proceedings of the 16th Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution (INFOCOM '97)*, pp. 1405–1413, April 1997.

[13] C. K. Toh, *Ad Hoc Mobile Wireless Networks: Protocols and Systems*, Prentice Hall, New York, NY, USA, 2002.

[14] L. Chen and W. B. Heinzelman, "QoS-aware routing based on bandwidth estimation for mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 561–572, 2005.

[15] T. A. Ramrekha, V. N. Talooki, J. Rodriguez, and C. Politis, "Energy efficient and scalable routing protocol for extreme emergency ad hoc communications," *Mobile Networks and Applications*, vol. 17, no. 2, pp. 312–324, 2012.

[16] N. Nomikos, D. Skoutas, D. Vouyioukas, C. Verikoukis, and C. Skianis, "Capacity maximization through energy-aware multimode relaying," *Wireless Personal Communications*, vol. 74, no. 1, pp. 83–99, 2014.

[17] J. Gomez, A. T. Campbell, M. Naghshineh, and C. Bisdikian, "PARO: supporting dynamic power controlled routing in wireless ad hoc networks," *Wireless Networks*, vol. 9, no. 5, pp. 443–460, 2003.

[18] A. Lindgren and O. Schelen, "Infrastructured ad-hoc networks," in *Proceedings of the International Conference on Parallel Processing*, pp. 64–70.

[19] J. J. Edwards, J. D. Brown, and P. C. Mason, "Using covert timing channels for attack detection in MANETs," in *Proceedings of the Military Communications Conference (MILCOM '12)*, pp. 1–7, Orlando, Fla, USA, October 2012.

[20] D. Djenouri and N. Badache, "Dynamic source routing power-aware," *International Journal of Ad-Hoc and Ubiquitous Computing*, vol. 1, pp. 126–136, 2006.

[21] G. Chakrabarti and S. Kulkarni, "Load balancing and resource reservation in mobile ad hoc networks," *Ad Hoc Networks*, vol. 4, no. 2, pp. 186–203, 2006.

[22] S. K. Bisoyi and S. Sahu, "Performance analysis of dynamic MANET ondemand (DYMO) routing protocol," *International Journal of Computer & Communication Technology*, vol. 1, no. 2,3,4, pp. 338–346, 2012.

[23] IETF, "Generalized mobile ad hoc network (MANET) packet/message format," http://tools.ietf.org/html/rfc5444.

[24] IETF, "Internet Protocol (IP)," RFC 791, http://tools.ietf.org/html/rfc791.

[25] G. Koltsidas, F.-N. Pavlidou, K. Kuladinithi, A. Timm-Giel, and C. Gorg, "Investigating the performance of a multipath DYMO protocol for ad-hoc networks," in *Proceedings of the IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, Athens, Greece, September 2007.

[26] D. Singh, A. K. Maurya, and A. K. Sarje, "Comparative performance analysis of LANMAR, LAR1, DYMO and ZRP routing protocols in MANET using Random Waypoint Mobility Model," in *Proceedings of the 3rd International Conference on Electronics Computer Technology (ICECT '11)*, vol. 6, pp. 62–66, Kanyakumari, India, April 2011.

[27] C. Kretschmer, S. Ruhrup, and C. Schindelhauer, "DT-DYMO: delay-tolerant dynamic MANET on-demand routing," in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops (ICDCS '09)*, pp. 493–498, Montreal, Canada, June 2009.

[28] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile ComputIng*, T. Imielinski and H. Korth, Eds., vol. 353, pp. 153–181, Springer, New York, NY, USA, 1996.

[29] "The network simulator ns-2," http://www.isi.edu/nsnam/ns/.

[30] N. Zorba, C. Skianis, and C. Verikois, *Cross Layer Designs in WLAN Systems*, vol. 1&2, Troubador Publishing, Leicester, UK, 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 268284, 11 pages
http://dx.doi.org/10.1155/2014/268284

Research Article

A Cross-Layer Approach to Minimize the Energy Consumption in Wireless Sensor Networks

Luca Catarinucci, Riccardo Colella, Giuseppe Del Fiore, Luca Mainetti, Vincenzo Mighali, Luigi Patrono, and Maria Laura Stefanizzi

Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy

Correspondence should be addressed to Luigi Patrono; luigi.patrono@unisalento.it

Received 31 October 2013; Accepted 17 December 2013; Published 22 January 2014

Academic Editor: Christos Verikoukis

Energy efficiency represents one of the primary challenges in the development of wireless sensor networks (WSNs). Since communication is the most power consuming operation for a node, many current energy-efficient protocols are based on duty cycling mechanisms. However, most of these solutions are expensive from both the computational and the memory resources point of view and; therefore, they result in being hardly implementable on resources constrained devices, such as sensor nodes. This suggests to combine new communication protocols with hardware solutions able to further reduce the nodes' power consumption. In this work, a cross-layer solution, based on the combined use of a duty-cycling protocol and a new kind of active wake-up circuit, is presented and validated by using a test bed approach. The resulting solution significantly reduces idle listening periods by awakening the node only when a communication is detected. Specifically, an MAC scheduler manages the awakenings of a commercial power detector connected to the sensor node, and, if an actual communication is detected, it enables the radio transceiver. The effectiveness of the proposed cross-layer protocol has been thoroughly evaluated by means of tests carried out in an outdoor environment.

1. Introduction

Smart environments are expected to become the main actors of the Next Internet, which will be no longer seen as a means to connect people to services but to access the resources made available by small smart objects, first of all sensors and actuators, adopting the machine-to-machine (M2M) paradigm. This new vision of the Internet fits into the broader concept of the Internet of Things, according to which the everyday objects that surround us will become proactive actors of the global Internet, with the capability of generating and consuming information for advanced applications [1]. Among all the wireless technologies enabling the new vision of the Internet, wireless sensor networks (WSNs) are the ideal choice because sensor nodes are able to self-configure and self-organize. These characteristics make them useful to be deployed even in hostile environments in order to detect the environmental parameters (temperature, light, humidity, etc.) without human intervention. Then, exploiting the wireless channel and the multihop communication among nodes, the collected data are sent to a central processing point or are exploited by user-customized mash-up applications [2]. Other strengths of this technology are represented by the low cost of devices, their small size, and their low power consumption. These simple yet fundamental functionalities are of great interest for a plethora of applications, such as building automation, surveillance, military operations, healthcare, and logistics, just to mention a few of them. However, the management of power consumption is still one of the main problems that are slowing the widespread diffusion of WSNs. Indeed, sensor nodes are usually battery powered and deployed in large areas in which changing or replacing batteries is impractical or completely unfeasible. Therefore, minimizing the power consumption in a node is a primary issue to be considered, and the use of effective solutions for increasing the nodes lifetime is fundamental in many applicative scenarios.

Let us observe that the power consumption of nodes is negligible in data sensing and processing procedures. On the contrary, the data communication towards the central

processing system has a strong impact on the nodes battery. This issue has a twofold cause: on the one hand, the radio transceiver implies a higher power consumption compared to the other components of the embedded device; on the other hand, the communication phase is associated with phenomena such as collisions, overhearing (i.e., listening of messages addressed to another node), overemitting (i.e., transmission of data to a node that cannot receive them), and idle listening (i.e., listening to the channel in absence of communications), which substantially reduce the nodes battery. To address these problems, many works in the literature are focused on energy saving solutions, mainly focused on the media access control (MAC) layer. The main goal of these protocol solutions is to minimize the activity time of the radio transceiver by properly setting the nodes duty cycle. In this way, each node is able to switch its radio component between ON and OFF state according to a predefined scheduling. However, most of these solutions result in being hardly implementable on real embedded devices, since they are expensive from both the computational and the memory resources point of view [3]. These issues suggest to combine new MAC protocols with hardware solutions able to further reduce the node's power consumption [4, 5]. In this context, an increasing number of current works in the literature propose the use of a secondary low-power radio, called the "wake-up radio", able to monitor the channel and wake up the node only when a communication is detected. In such a way, nodes can remain asleep for most of the time and activate their main radio transceiver to receive data only when they receive a signal on the wake-up radio. This particular behavior allows to minimize the idle listening periods, and, consequently, the nodes' power consumption. Wake-up radios can be categorized as active and passive based on whether they use a power supply. Active wake-up radios require a continuous power supply, while passive systems harvest energy to power themselves from the wake-up radio signal transmitted by the sender. Since these last ones do not dissipate any energy from the battery, they operate over a shorter range of distances compared to active wake-up radios.

In this work, a cross-layer solution, based on the combined use of a duty-cycling protocol [6] and a new kind of active wake-up circuit, based on a very-low-consumption radio frequency (RF) power detector never adopted in the literature so far, is presented and validated. To the best of our knowledge, the proposed solution represents the first study presented in the literature on the integration of an active wake-up radio with an energy-efficient MAC protocol for WSNs. The resulting solution is able to substantially reduce power consumption and to extend sensor nodes lifetime by preventing unnecessary awakenings of nodes. Specifically, a commercial power detector has been connected to the sensor node and used as active wake-up radio. The power detector is an electronic integrated circuit able to provide an output signal linearly proportional to the input RF signal. To further reduce the overall power consumption, the activation of the secondary radio is controlled by the implemented scheduling schema. In more detail, during the network setup phase, each node communicates with the time interval chosen for the transmission of its packets. In such a way, neighboring nodes can properly set their wake-up times. This way, every node knows in advance when it can switch to sleep mode, because no transmissions are scheduled, and when it must be awake for receiving data. However, since a neighbor might not have data to transmit, a node switches on its radio transceiver to receive data only when the power detector detects a transmission. In order to evaluate the effectiveness of the proposed system, the selected power detector was firstly characterized through a test bed approach, and then a performance comparison between the proposed solution and the enhanced duty-cycling solution, reported in [7], was carried out by using real hardware devices. Obtained results show that the energy efficiency using the proposed cross-layer solution is much greater than when using the standard duty-cycling protocol.

The rest of the paper is organized as follows. Section 2 summarizes the state-of-the-art of MAC energy-efficient MAC solutions and radio wake-up technology for WSNs. The characterization of the power detector is described in Section 3, while the MAC scheduler is defined in Section 4. In Section 5 numerical results are discussed. Conclusions are drawn in Section 6.

2. Related Works

Most energy-efficient communication protocols for WSNS are based on duty-cycling mechanisms. Such protocols fit into three main categories: preamble sampling, scheduling, and hybrid approaches.

Preamble-sampling MAC protocols are based on the low-powerl (LPL) [8] technique, according to which nodes periodically wake up for a short duration to sample the channel. If the channel is idle, nodes go back to sleep immediately; otherwise, they keep listening until a data frame is received or a timeout occurs. The transmission of a packet is preceded by a preamble that is as long as the channel sampling interval, so as to ensure that all potential receivers can detect the communication and stay awake to receive the data. B-MAC [9] is an early example of LPL protocol. It uses unsynchronized duty cycling in order to reduce the idle listening. WiseMAC [10] is similar to B-MAC but further optimizes transmission by allowing all nodes to record the radio sample phase of their neighbors. The wake-up tone is sent just before the receiver wakes up, saving a greater amount of energy. X-MAC [11] was the first LPL protocol to use a strobe preamble (i.e., a sequence of short preambles). Such short preambles contain the address of the receiver, and, therefore, nontarget nodes can immediately go back to sleep when they receive a strobe for another node. Furthermore, X-MAC uses the gap between two packets to accommodate an early ACK. Some protocols, such as SpeckMAC-D [12] and MX-MAC [13], repeat an actual data packet as the preamble. However, using data packet as the short preamble packet increases the idle listening period.

Scheduling approaches, such as S-MAC [14], T-MAC [15], DW-MAC [16], and PW-MAC [17], reduce the node duty cycle exploiting the use of a MAC scheduler. In particular, in S-MAC [14] nodes are organized into clusters composed

of three periods: SYNC, DATA, and SLEEP. All nodes of the same cluster wake up at the beginning of the SYNC period to synchronize clocks with each other. Nodes with packets to send contend the channel during the DATA period, while nodes that are not involved in a communication return to sleep at the start of the SLEEP period. T-MAC [15] improves S-MAC by using an adaptive timer able to reduce the wake-up duration and introducing the Future Requests To Send (FRTS) policy. Specifically, it uses a special scheme to decide when a wakeup period can end (i.e., no activation event has occurred for a certain amount of time). This design aims to achieve optimal wake-up periods under various traffic loads. However, it introduces overhearing because a node has to stay awake also if it is not involved in data transmission. The scheduling algorithm in DW-MAC [16] integrates scheduling and access control to maintain a proportional one-to-one mapping function between a DATA period and the subsequent SLEEP period, which minimizes scheduling overhead while ensuring that data transmissions do not collide at their intended receivers. PW-MAC [17] protocol further improves both S-MAC and T-MAC exploiting a scheduler based on a pseudorandom algorithm, which allows the sender to predict the next wake-up time of the receiver node. Let us observe that all scheduled approaches must rely on a tight time synchronization procedure, which results in high overhead and significant power consumption even when there are no useful data to send.

Hybrid approaches, such as SCP-MAC [18] and AS-MAC [19], combine preamble sampling with scheduling techniques. In more detail, SCP-MAC [18] synchronizes the wake-up time of neighboring nodes so that only a short preamble is required to wake the receiver up. This protocol reduces the overall nodes power consumption but it is not able to avoid the overhearing problem. Such problem is successfully addressed by the AS-MAC [19] protocol, coordinating asynchronously the wakeup times of neighboring nodes. One of the main disadvantages of this protocol is the inefficiency in broadcast transmissions, since it has to transmit every packet once for each neighbor. In [7], an energy efficient MAC protocol based on an asynchronous scheduler is presented. According to this protocol, every node has the complete list of the transmission times of its neighbors and knows in advance when it can switch to sleep mode, because no transmissions are scheduled. Furthermore, to solve the clock drift problem, a node updates its neighbors' list every time it receives a new data packet. However, if a transmission is scheduled but no data have to be transmitted, both the sender and its neighbors wake up and remain active for the whole wake period, even if an actual transmission is not in progress. In the last years, several papers proposed cross-layer energy saving solutions for WSNs. In [20, 21] authors present an optimization design and evaluation of the Distributed Queuing (DQ) MAC protocol. Specifically, in [20] they describe a novel cross-layer fuzzy-rule-based scheduling algorithm, which allows packet transmissions to be scheduled taking into account the channel quality among body sensors. In [21], the potential benefits of DQ MAC in terms of energy efficiency per information bit under saturation conditions are analyzed.

Current duty-cycling protocols can only reduce but not eliminate idle listening, which remains the main source of power dissipation in sensor networks. An alternative approach suggests to use an additional low-power wake-up radio component able to listen to the channel when the node enters the sleep mode and to wake up the main radio transceiver when channel activity is detected. To gain a benefit in energy efficiency, the additional radio must be of lower power than the main data receiver. Several different low-power active wake-up radios have been proposed in the literature. In [22], a super-regenerative architecture with a 1.9 GHz Bulk Acoustic Wave (BAW) resonator is used to reduce the power consumption of the wake-up radio. This approach has been further optimized in [23]. In this work, a 65 μW wake-up receiver is created, using a 1.9 GHz BAW resonator matching network for RF signal filtering. In [24], a zero-bias Schottky diode envelope detector is used to receive a PWM signal. Using this signal, the address decoder generates the clocking signal necessary for the activation of the decoding circuit. A three-stage wake-up scheme is introduced in [25]. In this approach, a very low power (on the order of nW) always-on stage is used to trigger an intermediate higher power (on the order of μW) stage for wake-up signal verification. Only if the wake-up signal is confirmed the main transceiver is activated. Other approaches for active wake-up radios are described in [26, 27]. Although there are several hardware proposals for active wake-up radios, not many physical implementations or commercialized products are available. Furthermore, to the best of our knowledge, no cross-layer solution, based on the combined use of a duty-cycling protocol and an active wake-up circuit, has been previously presented in the literature.

3. Power Detector Enabling Radio Wakeup

In order to exploit the desired cross-layer approach and to reduce the WSN power consumption, WSN nodes provided with radio wake-up systems should be designed, realized, and validated. In this section, once the requirements for the hardware wake-up system are individuated, a solution is provided.

In particular, the radio wake-up system should be able to activate the node wireless interface only when a radio frequency (RF) signal is sent towards such a node. In such a way, even if in certain time periods the node is turned off and consequently does not waste power, the wake-up system must be permanently turned on in order to sense potential active WSN nodes. For such a reason, a power consumption appreciably lower than the node is the first requirement.

The second crucial requirement deals with the wake-up sensitivity, which represents the minimum RF power guaranteeing the proper functioning of the device. As the sensitivity is strongly linked with the maximum working range, in order not to introduce bottlenecks into the overall system, values comparable with that of the WSN node are strongly desired.

Some minor requirements, such as compactness in order to be easily integrated into the WSN node and cost

effectiveness in order to slightly impact upon the node total cost, must be satisfied as well.

In a first step, by taking into account all the requirements as a whole, it can be certainly deduced that an active solution must be preferred to a passive one. Indeed, in order to wake a WSN node up, a certain power is necessary, and, considering the low RF signal power emitted by a WSN node, a simple RF passive energy harvester used as wakeup would guarantee too short working distances. Vice versa, more complicated RF energy harvester systems provided with a DC-DC charge pump and the related capacitor, such as those presented in [28, 29], despite allowing longer working ranges, introduce latencies and asynchronism (due to the charging and discharging phases of the capacitor) which can be hardly managed in a WSN.

The proposed wake-up circuit is based on the use of an RF power meter, an active device commonly adopted to measure even very low RF signals. An important peculiarity of an RF power meter is that it is able to give a significant output voltage (as it is active) proportional to the incident RF power. Consequently, in the WSN context, such a signal can be used to generate a trigger to wake up the node.

Among the different devices available on the market, the Texas Instrument LMV221 [30] has been selected. Indeed, it works properly around 2.4 GHz (working band from 50 MHz to 3.5 GHz), it guarantees a reasonably good sensitivity (−45 dBm), its supply voltage of 3 V is compatible with that of many commercial WSN nodes, the supply current is of only 7.2 mA, and, finally, it is rather inexpensive.

For a practical usage, in the first developed prototypal version, the LMV221 evaluation board, named LM221EVAL [31], has been connected to a 2.4 GHz dipole-like antenna and adopted. In particular, in order to validate the proposed radio wake-up solution, the LMV221EVAL board has been used to drive the MB954 board, a WSN node developed by ST Microelectronics. This board is powered by a 3 V battery pack (which can be also used to power the power meter) and is equipped with a 32-bit ARM CortexTM-M3 microcontroller operating at a clock frequency up to 24 MHz and embedding 16 Kbytes of RAM and 256 Kbytes of eFlash as ROM. It integrates a 2.4 GHz wireless transceiver compliant with the IEEE 802.15.4 standard and a power amplifier. The radio transceiver needs a transmission current of 21 mA and a receive current of 19 mA. These values are increased by the consumption of the CPU during the node lifecycle: during active periods the CPU needs 7.5 mA, whereas when the radio transceiver is OFF, it uses only 3 mA. The mounted microcontroller is highly optimized to guarantee high performance at very low power consumption. But most importantly, the selected board is equipped with 24 highly configurable GPIOs. Consequently, the voltage output of the power meter can be straightforwardly connected to one of the GPIO ports configured for analog to digital conversion, and, depending on such a voltage value, a switching-on/off trigger can be generated and opportunely managed to smartly control the radio interface, as thoroughly described later on in the paper (Figure 1).

Before developing the cross-layer solution, an accurate characterization of the properties of the integrated device has been performed. For this purpose, a simple scenario, consisting of one sender and one receiver, has been considered. In particular, during the experimental campaign, the sender, with a standard configuration, has been statically positioned in the center of a soccer field and the receiver, connected to the wake up circuit, has been used to measure the output voltage produced by the power meter when a signal is detected. The experiment has been repeated several times increasing at each run the distance between the two nodes. The main results obtained in such test are reported in Figure 2. The curves clearly show that the measured voltage decreases as the distance increases. In particular, the analysis has shown that the integrated device is no longer able to detect a node's transmission when the distance between the two devices becomes greater than 35 meters.

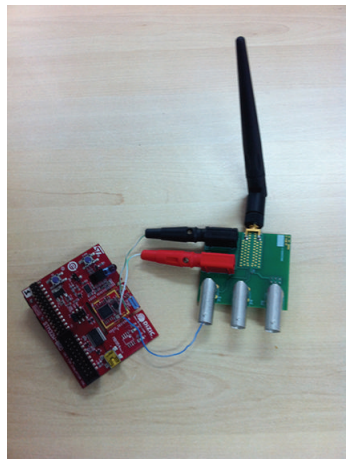


Figure 1: The MB954 evaluation board integrated with LM221-EVAL.

4. The Cross-Layer Radio Wake (CL-RW) Protocol

The basic idea of the defined protocol is to ensure smart awakenings; that is, nodes should wake up only when they actually have data to send or receive. In this perspective, during the network setup phase, each node chooses its transmission time, that is, the time instant at which it periodically can transmit; then, it communicates such information to its neighbors. In this way, in each duty cycle period, a node wakes up once to transmit and N times to receive, where N is the number of neighboring nodes. However, a node may not have data to transmit in a given period, and then the awakening of its neighbors would result in an unnecessary waste of energy. Just in these circumstances, the role of the power detector is fundamental: if the node has to wake up because the transmission of a neighbor is scheduled, the awakening occurs only if the wake-up device detects the presence of an effective communication.

For the sake of clearness, before describing the new scheduler in detail, some parameters used in the discussion are introduced below.

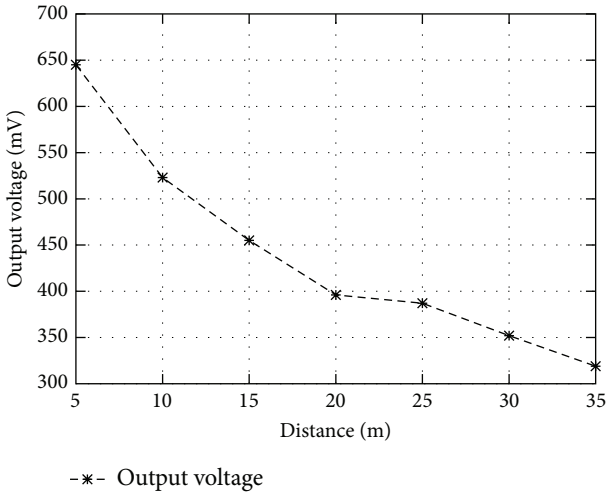


Figure 2: Output voltage versus the distance between the two devices.

(i) T_0 is the time interval (in seconds) between two subsequent transmissions. It is the same for every node and it is preconfigured.

(ii) Wake Time is the time interval (in seconds) in which a node can transmit the local buffered data or receive data from its neighbors.

(iii) Announce Packet (Pkt_{ANN}) is a signaling packet used by each node to advertise its presence; it contains the time interval between the current time and the next awakening time chosen for transmission.

(iv) Alert Packet ($\text{Pkt}_{\text{ALERT}}$) is a signaling packet used by a node to alert a neighbor about a possible collision.

(v) Full Packet (Pkt_{FULL}) is a signaling packet used by a node to inform its neighbors that it is out of the network.

(vi) Wake Packet (Pkt_{WAKE}) is a signaling packet used by a node to inform its neighbors that it is about to send data or a Pkt_{ANN}.

(vii) Wake-up Table (W_{TBL}) is a table used by each node to store information about the transmission times of its neighbors. Each table entry is associated with exactly one neighbor and contains the following information: (a) the ID of the neighbor, (b) the offset of the awakening time, and (c) the number of cycles of length T_0 during which no data have been received from the corresponding node.

In the following, the start-up phase and the periodic listening and sleep phase are described.

4.1. Network Startup. During the network initialization phase, all nodes stay awake for a certain time interval in order to detect the information useful to schedule their awakenings. In particular, they exchange information about their transmission time by sending Pkt_{ANN}s. On the reception of such a message from an unknown neighbor, the CL-RW MAC protocol updates its W_{TBL} by storing a new entry. However, before being stored, the information on the transmission time of the neighbor must be validated: the node verifies that the time chosen by the new neighbor does not overlap with the transmission intervals of the other neighboring nodes already stored into its W_{TBL}. If the verification procedure succeeds, the transmission time of the neighbor is converted into offset by subtracting an appropriate time interval and then it is stored according to the ascending order of the offsets. Otherwise, if the transmission interval chosen by the new node overlaps with any of the transmission intervals already in W_{TBL}, the node sends a $\text{Pkt}_{\text{ALERT}}$ to the new node, specifying the overlapping interval. In order to avoid collisions between packets, each node sends the $\text{Pkt}_{\text{ALERT}}$ after a waiting time, randomly chosen in a predefined time interval. In such a case, the new neighbor stores the received information into its W_{TBL} and it chooses a new transmission time. This mechanism also reduces one of the main problems that afflict ad hoc networks, that is, the hidden node problem: by leveraging the $\text{Pkt}_{\text{ALERT}}$, collisions among nodes two hops away are avoided. If the new node cannot find a valid transmission time, that is, the network is full, it communicates the information by broadcasting a Pkt_{FULL} and it turns off the radio. On the reception of such a message, all the neighbors, which have already stored an entry for that node, delete it.

Analyzing in more detail the transmission time selection procedure, we can say that each node chooses its own transmission time as a random value in a proper interval, also taking into account the choice done by its neighbors. This separation in time among transmissions of neighboring nodes leads to a reduced channel access contention. In more detail, if the W_{TBL} is empty, then the transmission time is randomly selected in the interval

$$\left[0, T_0 - (\text{WakeTime} + 2 * \text{TurnAroundTime})\right], \quad (1)$$

where WakeTime is the time window dedicated to data transmission and TurnAroundTime is the amount of time the radio needs for changing its state. If the W_{TBL} is not empty, then the node tries to set its own transmission time to a value different from those of its neighbors, in order to avoid collisions due to simultaneous transmissions. In particular, the node checks if there are two consecutive entries in the table, namely, ith and $(i + 1)$th, whose offsets difference is greater than

$$2 * \text{WakeTime} + 4 * \text{TurnAroundTime}. \quad (2)$$

If so, the transmission time is chosen within the interval

$$\left[\text{offset}\,[i] + D, \text{offset}\,[i + 1] - D\right], \quad (3)$$

where D = WakeTime + 2 * TurnAroundTime, whereas offset$[i]$ and offset$[i + 1]$ are the offsets associated with the ith and $(i + 1)$th entries, respectively. Note that the node also checks the time intervals:

$$\left[0, \text{offset}\,[0]\right], \qquad \left[\text{offset}\,[n], T_0 - D\right], \quad (4)$$

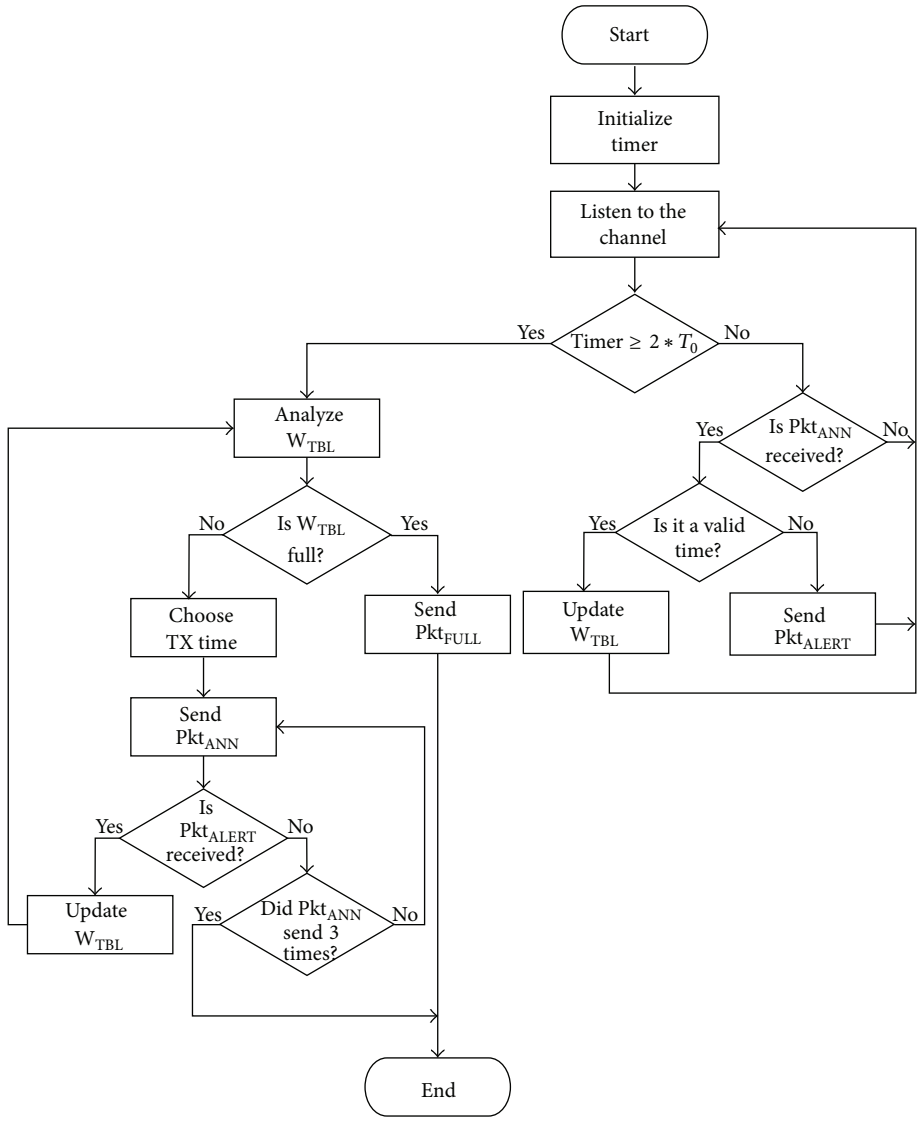


Figure 3: Flow chart of the network start-up phase.

where offset[0] and offset[n] are the offsets associated with the first and last entry, respectively.

In order to maximize the probability that all its neighbors receive the message, a node sends the Pkt_{ANN} three times. Figure 3 shows a simplified flow chart that clarifies the node behavior in the network start-up phase.

4.2. Steady State. After the start-up phase, the network enters in steady-state phase, during which two kinds of periodic events, namely, the transmission and the reception of packets, and one aperiodic event, that is, the arrival of a new node in the network, may happen. With regard to the periodic events, the node exploits the information stored in its W_{TBL}, by setting a timer for the next scheduled event. When the timer expires, if the event is a data transmission, the node checks the presence of packets in its queue. If there are buffered packets, then it sends a Pkt_{WAKE} to inform its neighbors about the imminent transmission. Otherwise, it keeps its radio transceiver OFF. When the transmission ends, the node waits for an ACK from the intended receiver, and, if no ACK is received, the message is sent again. At the end of its transmission interval, the node schedules the next event of the W_{TBL} and it switches off its radio transceiver. When the scheduled event is a data reception, the node activates the power detector in order to control if there is an incoming transmission; that is, the intended neighbor is sending a Pkt_{WAKE}. If so, it enables its radio transceiver, receives the data packet, and sends an ACK. On the contrary, if the power detector does not sense an incoming transmission until the end of a predefined timer, the node switches off the power detector and keeps its radio transceiver OFF. We can summarize the behavior of a node in the steady-state phase as a periodic transaction among the following five states:

(i) SLEEP-MODE: the node is inactive and waits for the next transmission or reception. In this state, the radio transceiver is OFF;

(ii) RX_WAKE: the node enters in this state when incoming transmission is scheduled and it verifies whether the transmission is occurring or not;

(iii) TX_WAKE: the node enters in this state when an its own transmission is scheduled and it verifies whether there are buffered data or not;

(iv) RX: in this state, the node is in its receiving time slot because the power detector has sensed an incoming transmission. Therefore, it waits for the data coming from the scheduled neighbor;

(v) TX: in this state, the node is in its transmission slot because it has verified that there are some data in its transmission buffer.

The state machine, reported in Figure 4, summarizes what we just said.

In order to accomplish the described behavior, both the transmission and the reception slots have two specific subintervals, namely, a checking subinterval and a communication sub-interval. For the transmission slot, the checking subinterval, called TX Wake Period, represents the time interval during which the node checks its queue for buffered packets, whereas the communication sub-interval, called TX Data Period, represents the time interval during which the node carries out the actual data transmission. Similarly, for the receiving slot, the checking sub-interval, called RX Wake Period, represents the time interval during which the node turns on the power detector to check the presence of an incoming transmission, whereas the communication subinterval, called RX Data Period, represents the time interval during which the node effectively receives data. Moreover, in order to correctly manage the arrival of new nodes in the network, both the transmission and the reception slots have another sub-interval, called Announcement Period. During this interval, the node enables the power detector in order to check whether a new node is announcing its presence or not. In the first case, the node turns on its radio component to receive the Pkt_{ANN}; otherwise it turns off the power detector and keeps its radio OFF until the start of the TX or RX Wake Period. The structure of the transmission and reception slots is shown in Figure 5. Figure 6, instead, shows the advantages resulting from the use of the power detector during the reception phase. While in the first duty cycle period Node 1 has some packets to transmit, during the other duty cycle intervals, it has no data in its buffer. In these situations, by leveraging the features of the power detector, Node 2 can keep its radio transceiver OFF, thus saving a considerable amount of energy.

As said, the proposed protocol is able to efficiently manage the entry of a new node in the network. In this situation, the new node first listens to the channel for a time interval equal to $2 * T_0$ with the aim of detecting the transmissions of its current neighbors, and then, for each packet received from an unknown node, it adds an entry in its W_{TBL}. Afterward, it exploits the Announcement Period of the transmission slots of its neighbors to communicate them the chosen transmission time, that is, to send the Pkt_{ANN}. In more detail, the new node sends a Pkt_{WAKE} in the first

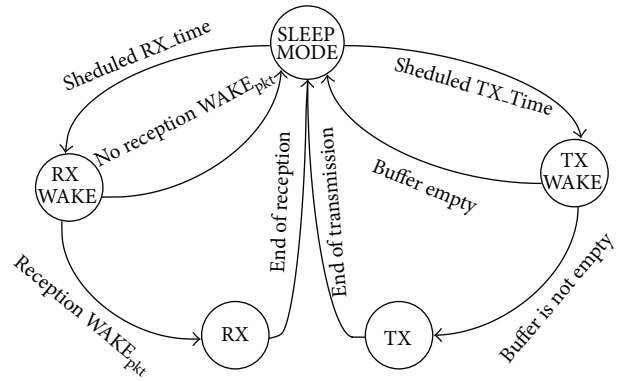


Figure 4: State machine of the steady-state phase.

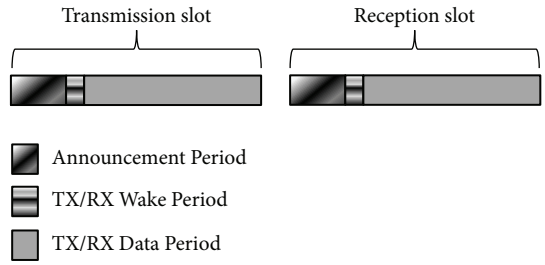


Figure 5: Structure of transmission and reception slots.

part of the Announcement Period of each neighbor to make sure that they sense it through the power detector. Once the neighbors receive Pkt_{WAKE}, they enable the radio component to receive the Announce Packet. If the transmission time stored in the Pkt_{ANN} does not overlap with the transmission time chosen by other nodes, the neighbors update their W_{TBL}. Otherwise, one or more nodes can communicate the bad choice by sending a $\text{Pkt}_{\text{ALERT}}$, as said in the previous section. Figure 7 summarizes the behavior just described. In the first two duty cycle periods, Node 3 listens to the transmissions of its neighbors and stores their transmission times in the W_{TBL}; then, in the third period, it sends its Pkt_{ANN} during the Announcement Periods of the two neighbors.

5. Results

The performance analysis of the proposed cross-layer solution was carried out by means of real test beds. This choice allowed us to evaluate the effectiveness of the proposed protocol as function of the hardware characteristics of both the board (e.g., clock speed, memory) and the wake-up circuit used. In more detail, a single-hop and a multihop scenario were considered in the tests. During the first experimental campaign (called STAR_TEST in the rest of the paper) a star topology, consisting of one receiver and four senders positioned in the same communication range, was considered. Instead, during the second test (called CHAIN_TEST in the rest of the paper), a chain network of five nodes was analyzed. All tests were carried out in an outdoor environment (i.e., a soccer field, without buildings in the surrounding area

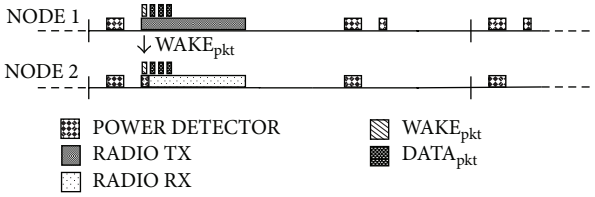


Figure 6: Reception phase.

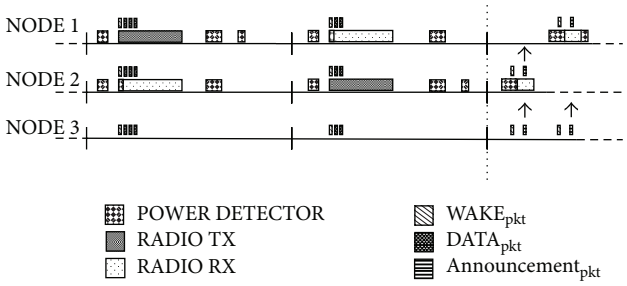


Figure 7: New node arrival.

Figure 8: Test bed at the University of Salento.

Table 1: Experimental parameters.

Parameter	Value
Network topology	Star, chain
Number of nodes	5
Number of packets	50
WakeTime	200 ms
Payload length	60 byte
Packet length (PHY layer)	91 byte
(Rate, T_0)	(1 packet every 10 seconds, 10) (1 packet every 30 seconds, 10) (1 packet every 60 seconds, 10) (1 packet every 120 seconds, 10)

as shown in Figure 8) inside the campus of the University of Silent, and, to limit the multipath problem due to the ground, the five MB954 used were positioned at a height of 1.5 m. In both network topologies, Node 1 was the sink and each node sent 50 packets towards the sink by adopting a Constant Packet Rate (CPR). Specifically, four different data rates were chosen: 1 packet every 10 seconds (high load), 1 packet every 30 seconds (medium load), 1 packet every 60 seconds (a typical data rate used in sensor networks [32]), and 1 packet every 120 seconds (very low load). Furthermore, to better appreciate the benefits derived by the use of the power meter, the proposed cross-layer solution was compared with the MAC solution implemented in [7] (called AS3-MAC in the rest of the section). The main idea of the AS3-MAC protocol is the concept of smart awake. In any duty cycle period, a node wakes up to both send and receive, but awakenings for reception are scheduled at the transmission times of the neighboring nodes. However, the awakenings (both in RX and TX state) are determined during the network initialization phase, their duration is fixed, and they remain unchanged during the steady state. In this way, the nodes wake up periodically to receive and to transmit even if there are no data to communicate. In both protocol solutions, the value of T_0 was set equal to 10 seconds, assuming that the running application can change its data rate without modifying the protocol layer settings. The main parameters of experimental campaigns are reported in Table 1. Let us observe that to evaluate the performance of the proposed solution without considering the routing traffic overhead, a static routing protocol was implemented.

In order to collect meaningful information, a custom data logging application was developed. The application, installed on the sink node, was able to send all received packets to a laptop working as a storage device. The data exchange between sink node and laptop was carried out by a serial communication. Each transmitted packet provides the information on the amount of time during which a node uses the radio transceiver and the wake-up circuit. In such a way, the overall node power consumption was measured. Finally, it is important to highlight that all tests were carried out by using the independent replications method and all results are characterized by a 95% confidence interval whose maximum relative error is 5%.

The performance results of the STAR_TEST are reported in Figure 9. The measured power consumption values are expressed in mW, whereas the four used data generation intervalS are labeled as DGI, indicating the elapsed time between two consecutive packet transmissions. It is important to observe that all reported power consumption values are evaluated by considering the activation periods of the main radio transceiver, the power meter, and the device microcontroller. In the considered network topology, all nodes consume the same energy. All nodes are in the same communication range, and, therefore, they have an equal number of neighbors, which determines the number of awakening in the W_{TBL}. The proposed cross-layer solution substantially outperforms the AS3-MAC protocol in terms of energy saving. This behavior can be noticed for all nodes by considering each data generation interval. In the graph, only one trend for the AS3-MAC solution is represented because the obtained results have shown that in this protocol the used data rate does not significantly affect the nodes' power consumption, since the idle power consumption is the dominating factor of the system power consumption. On the contrary, it is possible to note that in the proposed solution

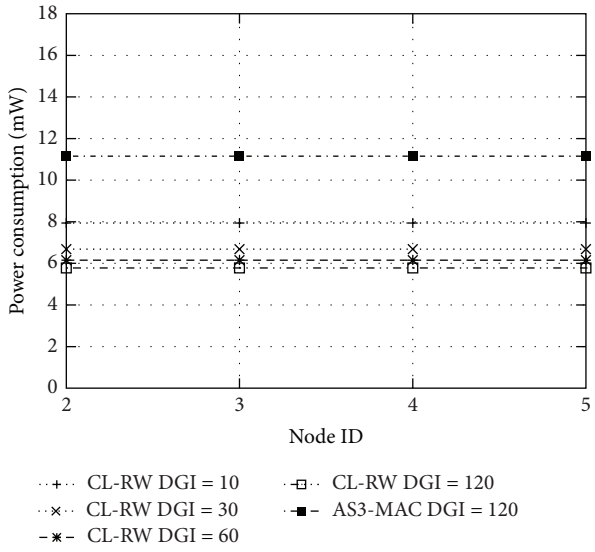


Figure 9: STAR_TEST: power consumption in the star topology.

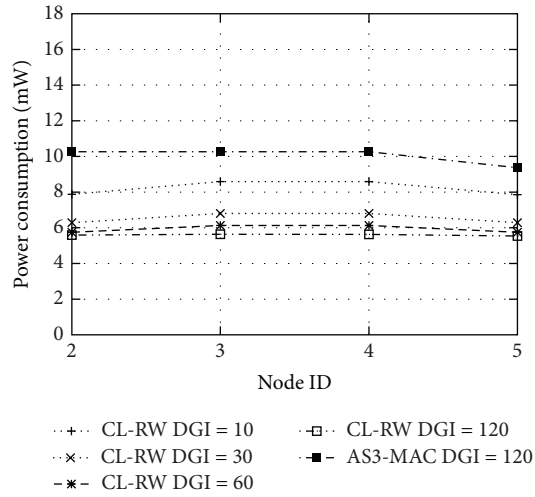


Figure 10: CHAIN_TEST: power consumption in the chain topology.

the power consumption behavior can be considered as a function of the data generation interval, since lower power consumption values are experienced at lower data rates. In this scenario, the energy saved by the proposed scheduler is around 48%, when lowest data rate is considered.

The results of the CHAIN_TEST are shown in Figure 10. As previously discussed, in both protocol solutions the transmission power consumption is not the dominating factor of the overall node power consumption. Therefore, the results do not show a significant difference among nodes closer to the sink, which forward messages generated by others too, and nodes further away. In the considered network topology, the farthest node shows lower power consumption due to the different number of neighbors. The last node in the chain has only one neighbor, and so it is awake for less time. It is important to observe that in the proposed cross-layer solution also Node 2 shows lower power consumption. According to our solution, a node turns on its main radio transceiver only when the power meter detects a packet transmission from a neighbor. Node 1 is the sink node and it does not perform packet transmissions during its activation periods. Therefore, in CL-RW protocol Node 2 does not turn on radio transceiver during transmission periods of the sink node. Furthermore, the curves in Figure 10 clearly show the linear relationship between the data rate and the nodes' power consumption, already discussed in the STAR_TEST results. Finally, obtained results confirm that the proposed solution outperforms the AS3-MAC protocol also using the chain topology. In particular, the energy saved by the proposed cross-layer protocol is about the 44%, when the lowest data rate is considered.

6. Conclusions

The reduction of the power consumption is one of the major issues in WSNs, as the lifetime duration is critical in this kind of networks. Among all the possible sources of energy waste, the communication phase and so the management of the radio transceiver are the most important issues to be addressed. In this work, a cross-layer approach based on the joint use of hardware and software solutions is proposed.

Firstly, a new kind of wake-up system for the node has been presented and validated. It is based on the integration between a commercial sensor node and a power meter circuit capable of switching (ON and OFF) the radio transceiver of the node according to the presence of an adequate RF signal. Then, a new duty-cycle-based communication protocol has been implemented, which exploits the power detector to activate, in each duty-cycle period, only the radio transceivers of those nodes actually involved in a communication. In such a way, the idle listening period is strongly reduced, and, consequently, the power consumption is reduced as well.

The proposed cross-layer solution has been deeply validated through a test bed approach aimed at a performance comparison with a similar MAC protocol already presented in the literature. The encouraging results presented and commented in the paper demonstrate the appropriateness of the proposed solution.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] L. Mainetti, L. Patrono, and A. Vilei, "Evolution of wireless sensor networks towards the internet of things: a survey," in *Proceedings of the 19th International Conference on Software, Telecommunications and Computer Networks (SoftCOM '11)*, pp. 16–21, September 2011.

[2] L. Mainetti, V. Mighali, S. L. Oliva, L. Patrono, and P. Rametta, "A novel architecture enabling the visual implementation of web of things applications," in *Proceedings of the 21th International*

Conference on Software, Telecommunications and Computer Networks (SoftCOM '13), 2013.

[3] D. Blasi, L. Mainetti, L. Patrono, and M. L. Stefanizzi, "Implementation and validation of a new protocol stack architecture for embedded systems," *Journal of Communication Software and Systems*, vol. 9, no. 3, pp. 157–169, 2013.

[4] L. Catarinucci, S. Guglielmi, L. Patrono, and L. Tarricone, "Switched-beam antenna for wireless sensor network nodes," *Progress in Electromagnetics Research C*, vol. 39, pp. 193–207, 2013.

[5] L. Catarinucci, S. Guglielmi, L. Mainetti, V. Mighali, L. Patrono, and M. L. Stefanizzi, "An energy-efficient MAC scheduler based on a switched-beam antenna for wireless sensor networks," *Journal of Communication Software and Systems*, vol. 9, no. 2, pp. 117–127, 2013.

[6] D. Alessandrelli, L. Patrono, G. Pellerano, M. Petracca, and M. L. Stefanizzi, "Implementation and validation of an energy-efficient MAC scheduler for WSNs by a test bed approach," in *Proceedings of the International Conference on Software, Telecommunications and Computer Networks (SoftCOM '12)*, pp. 1–6, 2012.

[7] D. Alessandrelli, L. Mainetti, L. Patrono, G. Pellerano, M. Petracca, and M. L. Stefanizzi, "Performance evaluation of an energy-efficient MAC scheduler by using a test bed approach," *Journal of Communication Software and Systems*, vol. 9, no. 1, pp. 84–96, 2013.

[8] C. J. Merlin and W. B. Heinzelman, "Duty cycle control for low-power-listening MAC protocols," *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1508–1521, 2010.

[9] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, pp. 95–107, November 2004.

[10] C. Enz, A. El-Hoiydi, J.-D. Decotignie, and V. Peiris, "WiseNET: an ultralow-power wireless sensor network solution," *Computer*, vol. 37, no. 8, pp. 62–70, 2004.

[11] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 307–320, November 2006.

[12] K.-J. Wong and D. K. Arvind, "SpeckMAC: low-power decentralised MAC protocols for low data rate transmissions in specknets," in *Proceedings of the 2nd International Workshop on Multi-hop Ad Hoc Networks: From Theory to Reality*, pp. 71–78, Florence, Italy, May 2006.

[13] C. J. Merlin and W. B. Heinzelman, "Network-aware adaptation of MAC scheduling for wireless sensor networks," in *Proceedings of the 3rd IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS '07) Poster Session*, June 2007.

[14] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 3, pp. 493–506, 2004.

[15] T. van Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the 1st ACM Conference on Embedded Networked Sensor Systems*, pp. 171–180, November 2003.

[16] Y. Sun, S. Du, D. B. Johnson, and O. Gurewitz, "DW-MAC: A low latency, energy efficient demand-wakeup MAC protocol for wireless sensor networks," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '08)*, pp. 53–62, May 2008.

[17] L. Tang, Y. Sun, O. Gurewitz, and D. B. Johnson, "PW-MAC: an energy-efficient predictive-wakeup MAC protocol for wireless sensor networks," in *Proceedings of the 30th IEEE Conference on Computer Communications (INFOCOM '11)*, pp. 1305–1313, Shanghai, China, April 2011.

[18] W. Ye, F. Silva, and J. Heidemann, "Ultra-low duty cycle MAC with scheduled channel polling," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 321–334, November 2006.

[19] B. Jang, J. B. Lim, and M. L. Sichitiu, "AS-MAC: an asynchronous scheduled MAC protocol for wireless sensor networks," in *Proceedings of the 5th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS '08)*, pp. 434–441, Atlanta, Ga, USA, October 2008.

[20] B. Otal, L. Alonso, and C. Verikoukis, "Highly reliable energy-saving mac for wireless body sensor networks in healthcare systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 553–565, 2009.

[21] B. Otal, L. Alonso, and C. Verikoukis, "Energy-efficiency analysis of a distributed queuing medium access control protocol for biomedical wireless sensor networks in saturation conditions," *Sensors*, vol. 11, no. 2, pp. 1277–1296, 2011.

[22] B. Otis, Y. Chee, and J. Rabaey, "A 400 μW-RX, 1.6 mW-TX super-regenerative transceiver for wireless sensor networks," in *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC '05)*, pp. 396–606, San Francisco, Calif, USA, February 2005.

[23] N. Pletcher, S. Gambini, and J. Rabaey, "A 65 μW, 1.9 GHz RF to digital baseband wakeup receiver for wireless sensor nodes," in *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC '07)*, pp. 539–542, San Jose, Calif, USA, September 2007.

[24] P. Le-Huy and S. Roy, "Low-power 2.4 GHz wake-up radio for wireless sensor networks," in *Proceedings of the 4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communication (WiMob '08)*, pp. 13–18, Avignon, France, October 2008.

[25] S. von der Mark, R. Kamp, M. Huber, and G. Boeck, "Three stage wakeup scheme for sensor networks," in *Proceedings of the SBMO/IEEE MTT-S International Microwave and Optoelectronic Conference*, pp. 205–208, July 2005.

[26] J. Ansari, D. Pankin, and P. Mähönen, "Radio-triggered wake-ups with addressing capabilities for extremely low power sensor network applications," in *Proceedings of the 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, pp. 1–5, Cannes, France, September 2008.

[27] B. V. d. Doorn, W. Kavelaars, and K. Langendoen, "A prototype low-cost wakeup radio for the 868 MHz band," *International Journal on Sensor Networks*, vol. 5, pp. 22–32, 2009.

[28] D. de Donno, L. Catarinucci, and L. Tarricone, "An UHF RFID energy-harvesting system enhanced by a DC-DC charge pump in silicon-on-insulator technology," *IEEE Microwave and Wireless Components Letters*, vol. 23, pp. 315–317, 2013.

[29] L. Catarinucci, R. Colella, D. de Donno, and L. Tarricone, "Fully-passive devices for RFID smart sensing," in *Proceedings of the Antennas and Propagation Society International Symposium (APSURSI '13)*, pp. 1–2, Orlando, Fla, USA, July 2013.

[30] "LMV221Texas Instrument," http://www.ti.com/product/lmv-221.

[31] "LM221EVAL Evaluation Board," http://www.ti.com/tool/lmv-221eval.

[32] O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis, "Collection tree protocol," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09)*, pp. 1–14, Berkeley, Calif, USA, November 2009.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 314049, 11 pages
http://dx.doi.org/10.1155/2013/314049

Review Article

A Survey on Deployment Algorithms in Underwater Acoustic Sensor Networks

Guangjie Han,[1,2] Chenyu Zhang,[1] Lei Shu,[3] Ning Sun,[1] and Qingwu Li[1]

[1] *Department of Information & Communication Systems, Hohai University, Changzhou 213022, China*
[2] *Changzhou Key Laboratory of Sensor Networks and Environmental Sensing, Changzhou 213022, China*
[3] *Guangdong Provincial Key Lab of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China*

Correspondence should be addressed to Guangjie Han; hanguangjie@gmail.com

Received 31 October 2013; Accepted 22 November 2013

Academic Editor: Joel J. P. C. Rodrigues

Node deployment is one of the fundamental tasks for underwater acoustic sensor networks (UASNs) where the deployment strategy supports many fundamental network services, such as network topology control, routing, and boundary detection. Due to the complex deployment environment in three-dimensional (3D) space and unique characteristics of underwater acoustic channel, many factors need to be considered specifically during the deployment of UASNs. Thus, deployment issues in UASNs are significantly different from those of wireless sensor networks (WSNs). Node deployment for UASNs is an attractive research topic upon which a large number of algorithms have been proposed recently. This paper seeks to provide an overview of the most recent advances of deployment algorithms in UASNs while pointing out the open issues. In this paper, the deployment algorithms are classified into three categories based on the mobility of sensor nodes, namely, (I) static deployment, (II) self-adjustment deployment, and (III) movement-assisted deployment. The differences of the representative algorithms in aspects of sensor node types, computation complexity, energy consumption, deployment objectives, and so forth, are discussed and investigated in detail.

1. Introduction

Recently, advances in wireless sensor networks (WSNs) have motivated the development of underwater acoustic sensor networks (UASNs), which have become a compelling technology to enable and enhance applications such as environment monitoring, resource exploration, disaster prevention, pollution detection, and military surveillance [1].

UASNs are composed of different kinds of sensor nodes (i.e., surface sink, underwater sensor nodes, etc.) to collaboratively perform monitoring tasks over a three-dimensional (3D) space. A three-dimensional UASN architecture is shown in Figure 1. UASNs consist of static sensor nodes which are deployed both on the water surface and underwater and automatic mobile sensor nodes to perform collaborative monitoring tasks over a given monitored space. Static sensor nodes usually consist of a sensing device, a microcontroller, and an acoustic transceiver with a limited amount of energy. Automatic mobile sensor nodes such as autonomous underwater vehicles (AUVs), unmanned underwater vehicles (UUVs), and low-power gliders typically have plenty of energy which can be supplemented when needed. According to the application requirements, different kinds of sensor nodes can be deployed in UASNs, that is, surface sinks, underwater nodes, bottom nodes, and automatic mobile nodes. Surface sinks are responsible for data collection and global position system (GPS) signal acquisition; surface sinks can be either stationary or mobile. Underwater nodes are equipped with floating buoys which can be inflated by pumps to adjust their depths to cover the entire monitored space. Usually, bottom nodes are anchored at the bottom of the ocean to monitor the two-dimensional (2D) area or collaborate with underwater nodes and automatic mobile nodes to fulfill monitoring tasks in 3D space. Automatic mobile nodes can receive GPS signals while floating on the ocean surface, and then dive to a fixed depth and move among underwater nodes following a predefined trajectory to help with localization or information gathering, and so forth. Events are detected by

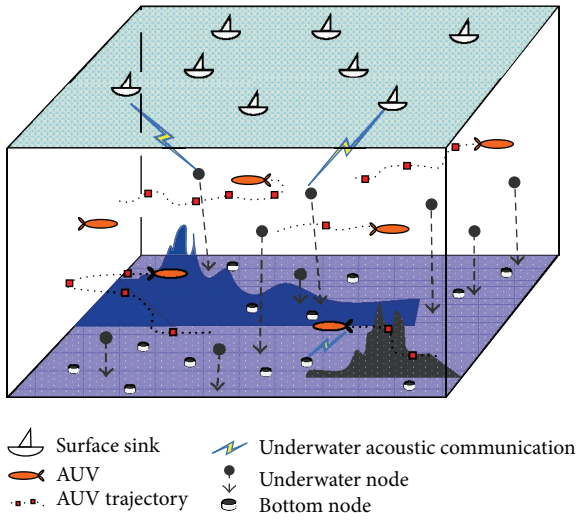


Figure 1: A three-dimensional UASN architecture.

sensor nodes locally and information is transferred to surface sinks by multihops or automatic mobile nodes using acoustic communication. Then, the data can be forwarded to onshore control centers which have larger storage capacity for future processing [2, 3].

In UASNs, sensor nodes communicate with each other via acoustic signals, while due to the unique characteristics of underwater acoustic channels (large propagation delay, high error rate, multipath effects, etc. [4]), new algorithms and protocols should be specifically designed for 3D UASNs. Besides, UASNs normally operate in uncertain and mobile environments where free-floating underwater sensor nodes drift slowly with the water current. As a result, the relative motion of the transmitter or receiver may create the Doppler effect. Moreover, UASNs are energy-limited. Energy supplement is difficult because it requires underwater vehicles which are costly to operate. In summary, node deployment algorithms designed for UASNs need to address the adverse physical channel conditions and water mobility while staying energy-efficient [5]. Generally, the deployment strategies support many fundamental network services, such as network topology control, routing, and boundary detection, which will further influence the network performance. Therefore, node deployment is one of the fundamental tasks in UASNs.

In this paper, we give an overview of the state of the art related to deployment algorithms in UASNs and categorize them into static deployment, self-adjustment deployment, and movement-assisted deployment. The differences of the representative algorithms in aspects of sensor node types, computation complexity, energy consumption, deployment objectives, and so forth, are discussed. Furthermore, the characteristics of deployment algorithms in three categories are investigated and compared in detail.

The remainder of this paper is organized as follows. Section 2 discusses design considerations, classification, and evaluation criteria of deployment algorithms in UASNs. Section 3 presents a detailed analysis of recent deployment algorithms in three categories, respectively, and summarizes the algorithms. Section 4 makes a conclusion and discusses future research issues of deployment algorithms in UASNs.

2. Design Considerations and Classification of Deployment Algorithms

2.1. Design Considerations for UASNs. Notice that most existing algorithms and protocols in WSNs are aiming at 2D sensor networks. For example, Capone et al. considered a heterogeneous network scenario and presented an optimal framework based on integer linear programming to locate wireless gateways [6]. Y.-R. Tsai and Y.-J. Tsai proposed a step-by-step node deployment algorithm aimed at minimizing location estimation of the entire large-scale WSNs [7]. Wang et al. transformed traffic-aware relay node deployment problem into a Euclidean Steiner Minimum Tree (ESMT) problem and proposed a hybrid algorithm to maximize lifetime for data collection [8]. Guerriero et al. mathematically proposed and defined different optimization deployment models to achieve high performance in terms of energy consumption and travelled distance [9]. However, these algorithms may no longer be effective in UASNs. Because three-dimensional networks require additional design and computation complexity, many problems cannot be solved by extension of two-dimensional algorithms. Besides, the design of 3D algorithms is more difficult than that of 2D. Thus, new algorithms should be specifically designed for 3D UASNs by exploring rich geometric properties of 3D UASNs. UASNs are different from WSNs in many ways [1, 3, 9].

High Latency. GPS signal cannot propagate through water and radio frequency (RF) signal can be absorbed by water. Propagation delay of underwater acoustic signal is five orders of magnitude higher than in RF terrestrial channels.

Limited Bandwidth and High Transmission Loss. Underwater acoustic channel has characteristics of a limited bandwidth and multipath fading, which result in high bit error rates.

Node Mobility. Positions of underwater sensor nodes are easily affected by water current or fish swarm.

Limited Energy. Battery power is limited and difficult to be recharged without utilizing solar energy. Besides, more complex signal processing consumes more energy.

High Cost. Underwater sensor nodes are easier to fail because of fouling and corrosion, so they need extra protective shell.

Sparse and 3D Deployment. Monitoring an ocean column requires a 3D deployment. The costly underwater sensor nodes make it more likely to be a sparse deployment.

Thus, the deployment of UASNs is deemed to be sparser than that of WSNs and more difficult to guarantee the network performance. In face of these characteristics, new algorithms and protocols for 3D UASNs should be specifically designed to achieve the optimal deployment of each sensor.

Many researchers are currently engaged in designing deployment algorithms for UASNs. Due to the wide range

of applications of UASNs and variety of underwater environment, different objective-oriented deployment algorithms could be designed to meet certain requirements. Typically, there are two types of architectures in UASNs, namely, (I) 2D UASNs for underwater bottom or surface monitoring and (II) 3D UASNs for underwater column monitoring [10]. In 2D UASNs, sensor nodes are anchored at the bottom of the ocean or floating on the ocean surface. In 3D UASNs, depths of sensor nodes can be adjusted by means of techniques.

2.2. Classification Based on Sensor Nodes' Mobility. There have been a large number of researches focusing on deployment issues in UASNs over the last few years. According to the mobility of sensors, the deployment algorithms can be classified into three categories, namely, static deployment, self-adjustment deployment, and movement-assisted deployment as shown in Figure 2. (I) Static deployment: all the sensors are static after initial deployment. Sensors are attached to surface buoys or anchored at the bottom of the ocean and they are assumed to have fixed positions. Static deployment is further classified into random deployment and regular deployment. (II) Self-adjustment deployment: underwater sensor nodes can adjust their depths by inflating floating buoys automatically or be driven to some desirable positions by mobile sensor nodes after initial deployment to meet certain requirements. Self-adjustment deployment is further classified into uniform coverage deployment and nonuniform deployment. (III) Movement-assisted deployment: there are underwater mobile sensor nodes patrolling over the monitored region to cooperate with other sensors to fulfill monitoring tasks. Some of the self-adjustment deployment algorithms and movement-assisted algorithms take sensor nodes' mobility into consideration, which is caused by water current or other marine animals.

2.3. Evaluation Criteria. We summarize typical deployment algorithms of each category in the following aspects.

Sensor Node Types. Typically, there are at least two kinds of sensors in UASNs, namely, sink node and underwater sensor nodes. Some researchers define other kinds of sensor node, such as surface gateway, AUV, and mobile data collector, according to their deployment algorithms.

Distributed or Centralized. Generally speaking, the distributed algorithms are more suitable for UASNs, due to their scalability and computational efficiency. The centralized algorithms are easier to implement, but may suffer from network attack which will result in network failure.

Computation Complexity. The computation complexity of an algorithm is of vital importance since it directly influences the execution time of the algorithm. Besides, the computation complexity also has an impact on energy consumption.

Energy Consumption. The total energy consumption of a network is determined by summing up energy consumption of all the sensor nodes in the network. Due to the limited energy resource, the algorithm should be energy efficient.

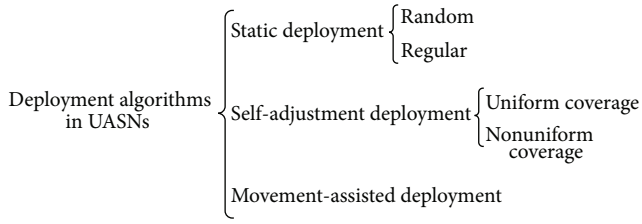


Figure 2: Classification of deployment algorithms for UASNs.

Deployment Objectives. Since sensor network is a kind of application-oriented network, deployment algorithms with different objectives could be designed to meet certain requirements.

Main Advantages. Although different deployment algorithms have their own characteristics, each algorithm has its own advantages.

3. Analysis of Deployment Algorithms

3.1. Static Deployment. For simplicity, many deployment algorithms assume that sensors are static after initial deployment in UASNs, with the consideration of practical reasons such as deployment cost and algorithm complexity. For 2D UASNs, sensors are usually deployed only on the water surface or at the bottom of the monitored region. Bottom sensors may be organized in a cluster-based architecture and forward collected data to surface sinks by multihop paths. For 3D UASNs, sensors are floating in different depths to observe the entire monitored region. In both architectures, sensors do not actively change their positions after initial deployment.

3.1.1. Random. Random deployment is the most practical way in deploying sensors. If no prior knowledge of the to-be-monitored region is available or deterministic deployment of sensors is very risky or infeasible, random deployment often becomes the only option. Usually, random deployment serves as an initial phase of self-adjustment and movement-assisted deployment strategies. Senouci et al. categorized random placement strategies into simple and compound [11]. Simple strategies are mere variants of the simple diffusion strategy, whereas compound strategies are realized by repeated simple diffusion. Through simulations, they give design guidelines in using stochastic deployment strategies.

3.1.2. Regular. The main characteristic of regular deployment is that sensors are anchored at the vertex of polygons or polyhedrons.

Pompili et al. provided a mathematical deployment analysis for both two-dimensional and three-dimensional architectures [12]. For two-dimensional architecture, they proposed a triangular-grid deployment algorithm to use the minimum number of sensors to achieve both optimal sensing

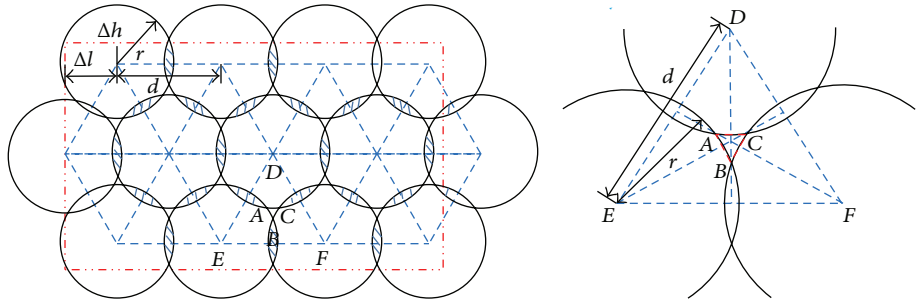


Figure 3: Triangular-grid deployment: (a) grid structure and side margins and (b) uncovered area (ΔABC).

and communication coverage. In triangular-grid deployment, sensors are deployed at the vertex of equilateral triangles. They mathematically proved that overlapping areas of neighboring sensors can be minimized and the coverage ratio is equal to 1, when length of equilateral triangle d is $\sqrt{3}$ times of sensing range r, as shown in Figure 3. There is a trade-off between the number of sensors and achievable sensing coverage. The authors studied the dynamics of a sinking object and evaluated the trajectory of a sinking object under the impact of water current, which can be used to estimate the final position of a sinking object.

Ibrahim et al. formulated the gateway optimal deployment issue as an integer linear programming (ILP) problem [13]. Later, they proposed a multiple surface-level gateways deployment algorithm based on heuristic approaches to enhance network performance [14]. They used greed algorithm and greed-interchange algorithm to select optimal gateway positions among $m \times m$ candidate locations which are mesh of points. The authors provided a multisink architecture and presented guidelines for deciding the number and positions of surface gateways in UASNs. However, it is a two-dimensional sensor deployment issue indeed.

Nazrul Alam and Haas aimed at finding a node deployment strategy with 100% sensing coverage of a 3D space, while minimizing the number of sensors needed [15]. In this paper, the authors defined a metric called volumetric quotient, which is the ratio of the volume of a polyhedron to the volume of its circumsphere. The higher the volumetric quotient, the smaller the number of nodes required for full 3D coverage. The paper gives a detailed analysis of several space-filling polyhedrons deployment, namely, cube deployment, hexagonal prism deployment, rhombic dodecahedron deployment, and truncated octahedron deployment as shown in Figure 4. Simulation results indicated that truncated octahedral cells result in the best strategy, and the higher the volumetric quotient is, the smaller the number of nodes that is needed to meet the coverage and connectivity requirements.

Felemban et al. formulated the optimal node placement as a nonlinear mathematical program with the objective of minimizing the transmission loss under a given monitored volume and number of sensor nodes [16]. Relay nodes are deployed next to each other in the 3D space forming tiled truncated octahedrons as presented in [15]. The main difference is that they considered the characteristics of underwater acoustic channels. Simulations showed that the operating frequency affects the number of nodes needed to cover a definite volume.

A UASN deployment algorithm (UDA) is proposed by Liu [17]. Different from [15, 16], UDA aims at determining and selecting the best cluster shape after partitioning the monitored region into layers and clusters. Clusters have shapes of cuboids, hexagonal prisms, rhombic dodecahedrons, and truncated octahedrons. Only the cluster-head in every cluster keeps alive for sensing and communicating. Once the cluster-head depletes its energy, a new cluster-head within the cluster will wake up and replace the old one. UDA sets different node densities at different layers to help prolonging lifetime of sensors close to sinks and balance the network energy consumption. However, there is no description about how to keep time synchronization when waking up a new cluster-head to maintain full connectivity and full coverage.

A multipath virtual sink architecture for UASNs is presented in [18] to overcome adverse link conditions. Virtual sinks are deployed at vertex of monitoring surface. The network dynamically selects shortest paths to deliver data over more routes to increase the probability of successful delivery instead of retransmissions. The algorithm ensures that data delivery continues to function even when a part of the network is temporarily nonoperational.

These static deployment algorithms enable easy and energy efficient deployment strategies for UASNs. They do not need auxiliary mobile sensor nodes to participate in deployment periodically. However, static deployment also suffers from some weaknesses. For instance, they cannot deal with real time event processing in large-scale UASNs.

3.2. Self-Adjustment Deployment. If the sensors have the ability of adjusting their positions in underwater environment after initial deployment to meet certain requirements, the corresponding algorithm is classified as the self-adjustment deployment. Each sensor is attached to a floating buoy. Sensor nodes can adjust their depths by controlling the length of the wires [19]. The depth adjustment system enables the underwater sensors to be deployed with a desired topology which can improve network connectivity and communication link over the whole monitored region. Moreover,

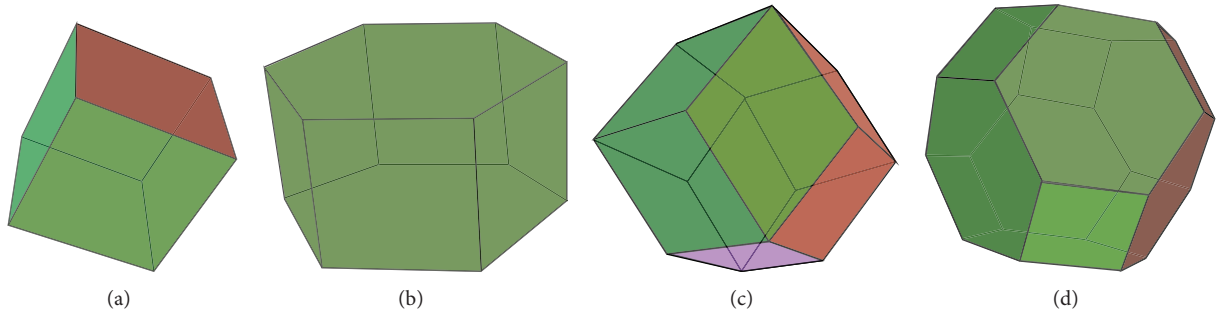


Figure 4: Several space-filling polyhedrons deployment: (a) cube deployment, (b) hexagonal prism deployment, (c) rhombic dodecahedron deployment, and (d) truncated octahedron deployment.

relocated sensors at different depths based on a local agreement can efficiently reduce the coverage overlaps among the neighboring nodes.

3.2.1. Uniform Coverage. For three-dimensional architecture in [12], Pompili et al. proposed three deployment strategies, namely, 3D-random, bottom-random, and bottom-grid. Sensors are initially deployed at the bottom of the ocean. Then, each sensor is assigned to a target depth. Sensors float to the desired positions to achieve the target coverage ratio. However, they neglected to describe how to adjust sensor depths to maximize coverage ratio in a 3D space.

Akkaya and Newell proposed a distributed node deployment algorithm to reduce coverage overlaps based on the idea of graph coloring by using the relocation ability of underwater sensor nodes in a vertical direction [20]. At first, sensors are randomly deployed at the bottom of the ocean. The new locations in the vertical direction are computed by cluster headers based on the group IDs of the sensors. Then, sensors continue to adjust their depths by applying repelling forces to minimize coverage overlaps until there are no sensing overlaps for a node or a maximum number of rounds have arrived as depicted in Figure 5. The authors also assessed the performance of the proposed algorithm under a UASN water current mobility model and gave a detailed analysis of simulation results.

Xia et al. introduced a rigid theory and defined rigidity-coverage value as the evaluation criteria for the positions of underwater sensors [21]. They established a novel underwater self-organization deployment mechanism based on rigidity-driven mobile strategy. Sensors move randomly in the ocean until they meet their neighboring nodes. When a sensor node s_c detects its neighboring node s_i, s_c and s_i will assemble in a crowd. It is a distributed deployment algorithm. Sensors execute the rigidity-driven deployment process periodically without time synchronization. The simulations were carried out in the areas of 10×10 (2D) and $10 \times 10 \times 10$ (3D) with the number of sensors varying from 10 to 15, respectively. Although they tested the performance of the deployment algorithm under the impact of water current, both the simulation area and the number of sensors are too small.

The mobility of underwater sensors makes the network topology slowly change and be inconveniently controlled. References [22, 23] aim at improving network coverage ratio and maintaining network connectivity when there are noncoverage areas or shadow zones. In [22], a dividing cube method is used to calculate the coverage ratio of an area. Then, two redeployment algorithms are proposed, of which one is based on the idea of adding new nodes, whereas the other is by the means of moving some redundant nodes from overcovered areas to noncoverage areas. Domingo presented an adaptive topology reorganization scheme to minimize the transmission loss and maintain network connectivity when a UASN is affected by shadow zones [23]. Sensor nodes are double units, operating as a single sensor, which are decoupled into two sensor nodes in the presence of a shadow zone. For instance, the sensor node S_i is uncoupled into two sensor nodes $S_{i\to 1}$ and $S_{i\to 2}$ under the presence of a shadow zone as shown in Figure 6. $S_{i\to 1}$ and $S_{i\to 2}$ are connected to each other by a wire to maintain robust communication. In this way, connectivity between two neighbor nodes is maintained and latency as well as power consumption is reduced.

3.2.2. Nonuniform Coverage. Since a UASN is a monitoring network, event detection is one of its main tasks. In nonuniform coverage deployment, how to adjust the positions of sensors according to the dynamic environment and targets to cover the event area in an optimal manner is a key issue in UASNs.

By simulating the behaviors of fish swarm and considering congestion control, a fish swarm-inspired sensor deployment (FSSD) algorithm was presented [24]. Sensors are nonuniformly deployed in the monitored region. FSSD algorithm enables sensors to cover the event area automatically by executing prey, follow, and swarm processes. FSSD algorithm seeks to make node distribution density match with event distribution density. They defined an evaluation criterion called event set coverage efficiency to evaluate the deployment performance. FSSD algorithm is a distributed algorithm with low computation complexity and high energy consumption.

The monitored area is divided into sectors with varying geographic size and acoustic characteristics [25]. Game Theory Field Design (GTFD) model is used to allocate sensors to sectors according to the visitation probabilities

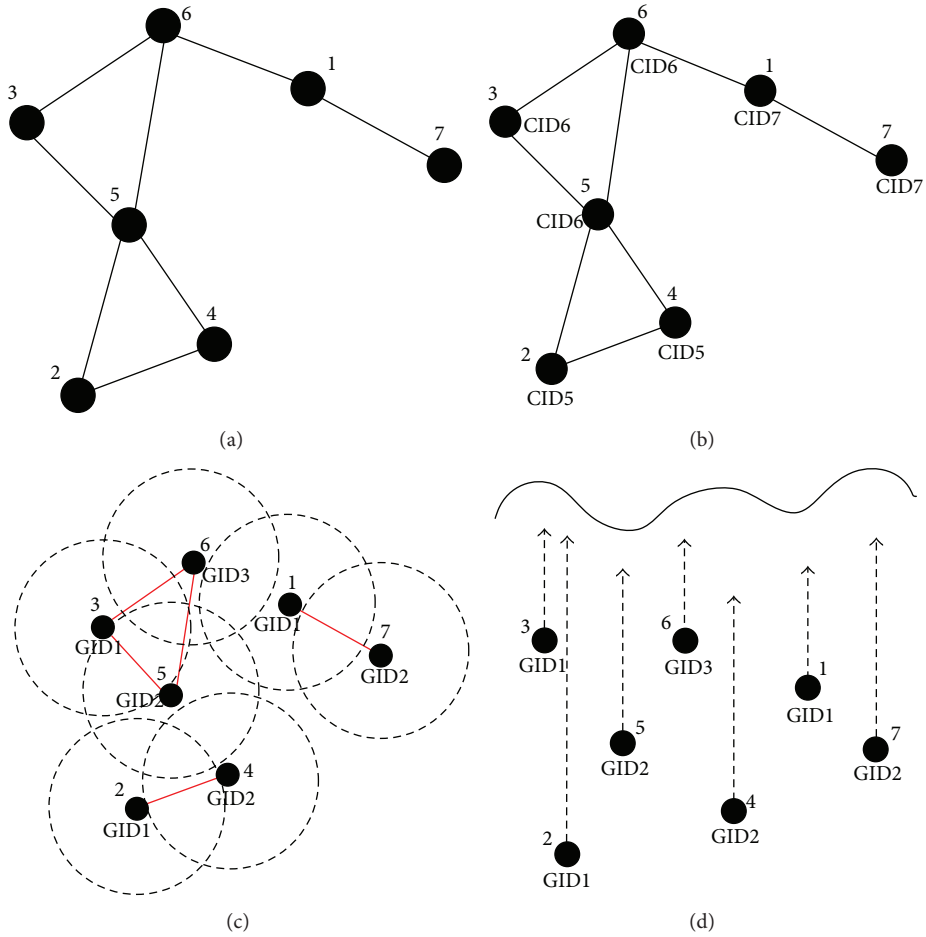


Figure 5: Distributed topology adjustment: (a) initialization, (b) clustering, (c) grouping, and (d) depth adjustment.

of an adversary. The authors compared GTFD model with Size-Aware Field Design (SAFD) model and Radius-Aware Field Design (RAFD) model. SAFD and RAFD only consider either the size of a sector or acoustic characteristics, respectively. Simulations showed that GTFD offers a significant improvement in terms of overall field detection capability against intelligent adversaries. But they did not explain how to deploy sensors in each sector. Besides, it is a centralized algorithm with 2D architecture and has high computation complexity.

Aitsaadi et al. addressed the issue of deploying a UASN in an area characterized by the geographical irregularity of the sensed event [26]. They proposed a geometric method called Differentiated Deployment Algorithm (DDA) based on a mesh representation method. By using a multiobjectives nonlinear optimization method, the optimal solution of DDA could be obtained. In DDA, authors considered triangles and rectangles during the hierarchical mesh. The basic idea is to permit progressive meshes division of a sensor field area as long as it is considered beneficial. New sensor nodes could be placed in equidistant locations to the vertices of the divided mesh. Or place an added node in the mid of one given arc of the mesh. It is a centralized algorithm, which is only applicable to static monitoring environment.

3.3. Movement-Assisted Deployment. In movement-assisted deployment, there exists AUVs, UUVs, low-power gliders, or other kinds of underwater mobile sensor nodes. They patrol over the monitored region under predefined trajectory to cooperate with other sensor nodes to fulfill monitoring tasks. Sensors can be attached to mobile sensor nodes as well and they can be driven to some desirable positions after initial deployment. Since underwater sensor nodes are fairly expensive and the costs quickly rise for deep water, underwater mobile sensor nodes can play more important roles compared with ordinary sensors in data collection, maximizing network coverage and real time event processing with limited hardware [27].

Teixeira et al. aimed at finding a control strategy that is able to drive a formation of AUVs from the initial to the target set of positions under the effect of external disturbances such as ocean current within a specified time [28]. They presented a leader-follower solution that relies on a simple uncertainty model to trigger surfacing events and two control strategies to position the agent within a certain distance of the target in the presence of disturbances.

Liu et al. and Ibrahim et al. studied the benefits of dynamic gateway redeployment [29, 30]. Both of the redeployment algorithms are triggered at fixed intervals. Liu et al.

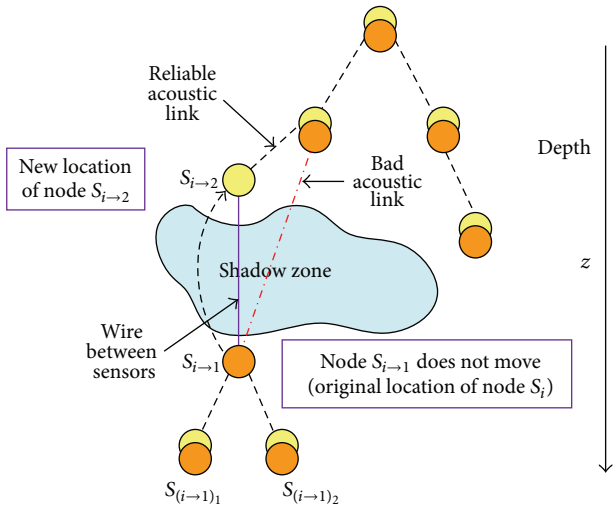


Figure 6: New scenario in the 3D UASN with a shadow zone.

proposed a prediction assisted dynamic surface gateway placement (PADP) algorithm for mobile underwater sensor networks [29]. PADP applies a tracking scheme called interacting multiple model (IMM) to predict sensors positions, uses branch-and-cut method to solve optimization problem, and adopts a disjoint-set data structure to partition sensors into clusters. By using PADA, gateways are deployed based on both sensor nodes current and future positions of sensor nodes, so coverage is well maintained. Ibrahim et al. modeled surface gateway deployment problem as a graph optimization problem [30]. The problem is to choose a subset of the candidate surface gateway satisfying a set of flow conservation constraints. The main difference from the gateway deployment problem they addressed in [14] is that they considered a dynamic gateway deployment strategy to cope with changes and fine-tune deployments. They also integrated the gateway deployment optimization framework into Aqua-Sim, a NS2-based UASN simulator.

The idea of mobile data collectors has been proposed in the literature for UASNs [31, 32]. Data collectors patrol over the monitored region under predefined trajectory through an optimization process. When the network is not a real time processing application, data can be stored at sensors until a mobile data collector is in the vicinity. Alsalih et al. divided the lifetime of the network into fixed length rounds and found the optimal locations of data collectors together with multihop routing paths using integer linear program (ILP) solver at the beginning of each round. The delay tolerant placement and routing (DTPR) scheme and the delay constrained placement and routing (DCPR) scheme are proposed to deliver the generated data from all sensor nodes to data collectors. The DCPR scheme has a boundary on how much time a data unit may spend on its way to a data collector. The objective function takes into account both the current residual energy and future energy expenditure of each sensor node. Experimental results showed that both of the schemes have the potential to prolong the lifetime of a UASN significantly.

The deployment of AUVs to collect data from underwater sensor networks has drawn much attention recently [33, 34]. AUV must plan its trajectory to minimize travel distance and maximize information gathering. Hollinger et al. proposed methods for solving this problem by extending approximation algorithms of variants of traveling salesperson problem (TSP) as shown in Figure 7 [33]. Williams developed a new adaptive strategy for performing data collection with a sonar-equipped AUV by adapting AUV survey route, which is based on the environmental characteristics observed in the data itself [34].

Yoon and Qiao focused on how to intelligently control multiple AUVs to complete the overall mission [35]. They proposed a cooperative rendezvous scheme called synchronization-based survey (SBS) to facilitate cooperation among several AUVs when surveying a large area. SBS has three variants, namely, alternating column synchronization (ACS), strict line synchronization (SLS), and x synchronization (XS). In SBS, AUVs form an intermittently connected network (ICN) in that they periodically meet each other for data aggregation, control signal dissemination, and AUV failure detection and recovery.

3.4. Summaries. In this section, we summaries the above-mentioned algorithms based on evaluation criteria presented in Section 2. From Table 1, we can conclude that each algorithm has its own advantages and no one is absolutely the best.

Generally speaking, most of the static deployment algorithms are centralized because sensor nodes are deployed uniformly at the beginning of the algorithms. Computation complexity and energy consumption is relatively low. A majority of static deployment algorithms aim at maximizing network coverage and minimizing sensor nodes that are needed in UASNs. However, the algorithms always neglect the mobility caused by water current which is inevitable in underwater environment.

In general, computational complexity and energy consumption of the self-adjustment deployment algorithms are larger than that of static deployment algorithms, because some of the sensor nodes can adjust their positions automatically to meet certain requirements. Due to unpredictable movement of sensor nodes in hybrid UASNs, some of the literature take sensor mobility into consideration by using a mobile model.

Since almost all the movement-assisted algorithms deploy mobile sensor nodes to help improving the network performance, the energy consumption of the movement-assisted algorithms is the largest among the three categories. Considering the costs and functions of mobile sensor nodes, fewer mobile sensor nodes are deployed in a UASN. Thus, the deployment problem is transformed into a path planning problem of mobile sensor. Most movement-assisted algorithms regard minimizing the total travel time and travel distance as one of their deployment objectives.

Table 1: Summary of the deployment algorithms.

Categories	Algorithms	Sensor types	Distributed or centralized	Computation complexity	Energy consumption	Deployment objectives	Main advantages
Static deployment	Ref. [12]	uw-sensor, UG	Centralized	Low	Low	Minimize overlapping coverage	Evaluate sinking objects' trajectory
	Ref. [14]	uw-sensor, UG	Distributed	High	Low	Gateway deployment optimization	Complexity analysis
	Ref. [15]	uw-sensor	Distributed	High	Low	Guarantee 100% coverage	Analysis of several space-filling polyhedrons
	Ref. [16]	uw-sensor, SG, RN	Centralized	High	Low	Minimize transmission loss	Detailed simulation analysis
	Ref. [17]	uw-sensor, CH	Centralized	High	Low	Maximize network lifetime	Clustering and duty-cycled network
	Ref. [18]	uw-sensor, virtual sink	Distributed	Low	Low	Ensure successful data delivery	Robustness
Self-adjustment deployment	Ref. [20]	uw-sensor, CH	Distributed	Low	High	Minimize overlapping coverage	Consider the influence of water current
	Ref. [21]	uw-sensor	Distributed	Low	High	Maximize coverage and connectivity	Consider the influence of water current
	Ref. [22]	uw-sensor	Centralized	Low	High	Minimize noncoverage areas	Coverage maintenance
	Ref. [23]	uw-sensor	Distributed	Low	Low	Maintain network connectivity	Energy-efficient
	Ref. [24]	uw-sensor	Distributed	Low	High	Cover the event areas automatically	Swarm intelligent algorithm
	Ref. [25]	uw-sensor	Centralized	High	Low	Allocates sensors follow event distribution	Event-driven-based nonuniform coverage
	Ref. [26]	uw-sensor	Centralized	High	Low	Achieve differentiated deployment	Differentiated coverage
Movement-assisted deployment	Ref. [28]	uw-sensor, AUV	Distributed	Low	High	Drive the AUV to target positions	Event-based
	Ref. [29]	uw-sensor, SG	Distributed	Low	High	Maximize coverage	Predict future positions
	Ref. [30]	uw-sensor, SG	Distributed	High	High	Minimize average end-to-end delay	Finding a near-optimal solution
	Ref. [31]	uw-sensor, DC	Distributed	High	High	Maximize network lifetime	Predict future energy expenditure
	Ref. [32]	uw-sensor, DC	Distributed	High	High	Maximize network lifetime	Delay constraints
	Ref. [33]	uw-sensor, AUV	Centralized	High	High	Minimize travel time	Consider different communication quality
	Ref. [34]	uw-sensor, AUV	Centralized	Low	High	Minimize travel time	Ensure the quality of collected data
	Ref. [35]	uw-sensor, AUV	Distributed	Low	High	Minimize travel time and distance	Failure detection and recovery

Uw-sensor: underwater sensor, UG: underwater gateway, SG: surface gateway, RN: relay node, CH: cluster head, SG: surface gateway, DC: data collector.

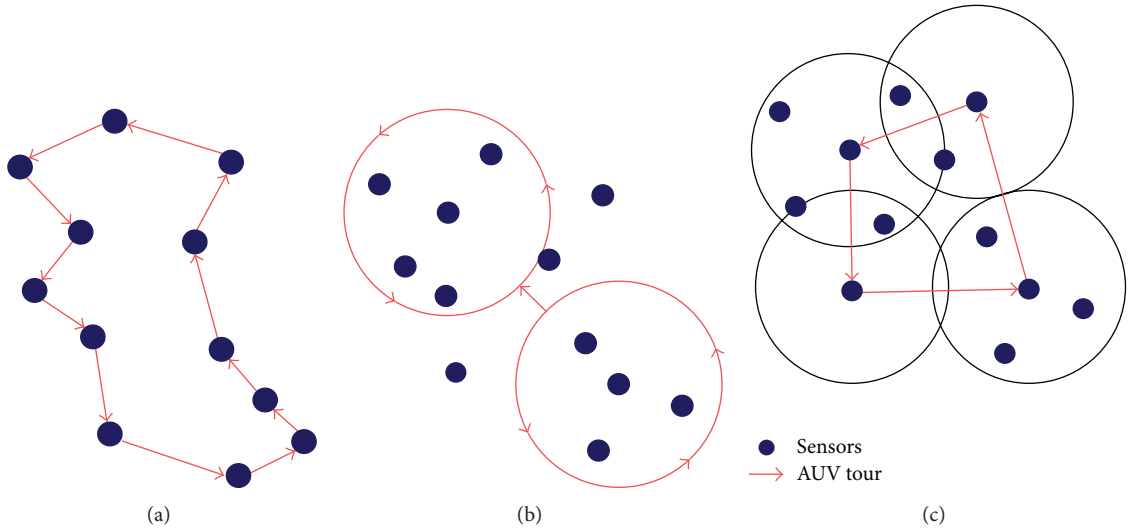


Figure 7: Example tours using different neighborhood types: (a) standard traveling salesperson tour, (b) tour circling a maximal independent set of neighborhoods, (c) tour visiting the center of a covering set of neighborhoods.

4. Conclusion and Future Research Issues

In this paper, we investigate the deployment issue in UASNs, which is a fundamental problem closely related to the quality of service of UASNs. We classify recent underwater deployment algorithms into three categories and give a comprehensive survey of them. Then, we summarize typical deployment algorithms in terms of sensor node types, computation complexity, energy consumption, and so forth, and make comparisons of three categories.

In recent years, many innovative solutions and ideas are used to design deployment algorithms in UASNs. The future research issues of deployment algorithms are possibly as follows.

(i) Since underwater sensor nodes inevitably move with water current, designing a mobility model is an important issue. When taking mobility into consideration, underwater deployment algorithms will become more effective.

(ii) The existing deployment algorithms are mainly suitable to small-scale networks. However, many practical underwater monitored spaces are large-scale. Thus, deployment algorithms which are suitable for large-scale UASNs should be specifically designed.

(iii) Since almost all the applications in UASNs are closely related to the locations of sensor nodes and the studying of node deployment also serves for localization, deployment algorithms and optimal path planning for mobile anchor nodes in large-scale UASNs should be specifically designed.

(iv) A majority of the current deployment algorithms are based on a credible environment. However, in real applications, sensor nodes may be deployed in an unsafe and complex environment. Thus, failure node recovering algorithms or redundant node deployment algorithms are needed in order to avoid network partition.

(v) To the best of our knowledge, few researches focus on the deployment of duty-cycled sensor nodes in UASNs to dynamically collaborate to prolong network lifetime. In duty-cycled environment, sensor nodes are allowed to be active or asleep periodically to reduce unnecessary energy consumption.

(vi) Since underwater sensor nodes are fairly expensive and the costs quickly rise for deep water, sensor nodes must have different sensing and communication ranges in real applications. Thus, future researches should focus on designing deployment algorithms for heterogeneous UASNs.

(vii) It is important to design systematic evaluation criteria to analyze the performance of different underwater deployment algorithms.

Acknowledgments

The work is supported by the Applied Basic Research Program of Changzhou Science and Technology Bureau, no. CJ20120028, the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, the Applied Basic Research Program of Nantong Science and Technology Bureau, no. BK2013032, and the Natural Science Foundation of Jiangsu Province of China, no. BK20131137.

References

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[2] H.-P. Tan, R. Diamant, W. K. G. Seah, and M. Waldmeyer, "A survey of techniques and challenges in underwater localization," *Ocean Engineering*, vol. 38, no. 14-15, pp. 1663–1676, 2011.

[3] R. B. Manjula and S. S. Manvi, "Issues in underwater acoustic sensor networks," *International Journal of Computer and Electrical Engineering*, vol. 3, no. 1, pp. 1793–8163, 2011.

[4] M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "Localization techniques for underwater acoustic sensor networks," *IEEE Communications Magazine*, vol. 48, no. 12, pp. 152–158, 2010.

[5] J. Partan, J. Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," in *Proceedings of the 1st ACM International Workshop on Underwater Networks (WUWNet '06)*, pp. 17–24, ACM, September 2006.

[6] A. Capone, M. Cesana, D. D. Donno, and I. Filippini, "Deploying multiple interconnected gateways in heterogeneous wireless sensor networks: an optimization approach," *Computer Communications*, vol. 33, no. 10, pp. 1151–1161, 2010.

[7] Y.-R. Tsai and Y.-J. Tsai, "Sub-optimal step-by-step node deployment algorithm for user localization in wireless sensor networks," in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC '08)*, pp. 114–121, June 2008.

[8] F. Wang, D. Wang, and J. Liu, "Traffic-aware relay node deployment: maximizing lifetime for data collection wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1415–1423, 2011.

[9] F. Guerriero, A. Violi, E. Natalizio, V. Loscri, and C. Costanzo, "Modelling and solving optimal placement problems in wireless sensor networks," *Applied Mathematical Modelling*, vol. 35, no. 1, pp. 230–241, 2011.

[10] J. Chen, E. Shen, and Y. Sun, "The deployment algorithms in wireless sensor net works: a survey," *Information Technology Journal*, vol. 8, no. 3, pp. 293–301, 2009.

[11] M. R. Senouci, A. Mellouk, and A. Aissani, "An analysis of intrinsic properties of stochastic node placement in sensor networks," in *Proceedings of the Global Communications Conference (GLOBECOM '12)*, pp. 494–499, December 2012.

[12] D. Pompili, T. Melodia, and I. F. Akyildiz, "Three-dimensional and two-dimensional deployment analysis for underwater acoustic sensor networks," *Ad Hoc Networks*, vol. 7, no. 4, pp. 778–790, 2009.

[13] S. Ibrahim, J.-H. Cui, and R. Ammar, "Surface-level gateway deployment for underwater sensor networks," in *Proceedings of the Military Communications Conference (MILCOM '07)*, pp. 1–7, October 2007.

[14] S. Ibrahim, J.-H. Cui, and R. Ammar, "Efficient surface gateway deployment for underwater sensor networks," in *Proceedings of the 13th IEEE Symposium on Computers and Communications (ISCC '08)*, pp. 1177–1182, July 2008.

[15] S. M. Nazrul Alam and Z. J. Haas, "Coverage and connectivity in three-dimensional networks," in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MOBICOM '06)*, pp. 346–357, September 2006.

[16] M. Felemban, B. Shihada, and K. Jamshaid, "Optimal node placement in underwater wireless sensor networks," in *Proceedings of the Advanced Information Networking and Applications (AINA '13)*, pp. 492–499, March 2013.

[17] L. Liu, "A deployment algorithm for underwater sensor networks in ocean environment," *Journal of Circuits, Systems and Computers*, vol. 20, no. 6, pp. 1051–1066, 2011.

[18] W. K. G. Seah and H.-X. Tan, "Multipath virtual sink architecture for underwater sensor networks," in *Proceedings of the OCEANS—Asia Pacific*, pp. 1–6, May 2007.

[19] D. Pompili, T. Melodia, and I. F. Akyildiz, "Deployment analysis in underwater acoustic wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Underwater Networks (WUWNet '06)*, pp. 48–55, September 2006.

[20] K. Akkaya and A. Newell, "Self-deployment of sensors for maximized coverage in underwater acoustic sensor networks," *Computer Communications*, vol. 32, no. 7–10, pp. 1233–1244, 2009.

[21] N. Xia, Y. Zheng, H. Du, C. Xu, and R. Zheng, "Rigidity driven underwater sensor self-organized deployment," *Chinese Journal of Computers*, vol. 36, no. 3, 2013.

[22] L. Bin, F. Ren, C. Lin, Y. Yang, R. Zeng, and H. Wen, "The redeployment issue in underwater sensor networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 1–6, December 2008.

[23] M. C. Domingo, "Optimal placement of wireless nodes in underwater wireless sensor networks with shadow zones," in *Proceedings of the 2nd IFIP Wireless Days (WD '09)*, pp. 1–6, December 2009.

[24] N. Xia, C.-S. Wang, R. Zheng, and J.-G. Jiang, "Fish swarm inspired underwater sensor deployment," *Acta Automatica Sinica*, vol. 38, no. 2, pp. 295–302, 2012.

[25] E. F. Golen, S. Mishra, and N. Shenoy, "An underwater sensor allocation scheme for a range dependent environment," *Computer Networks*, vol. 54, no. 3, pp. 404–415, 2010.

[26] N. Aitsaadi, N. Achir, K. Boussetta, and G. Pujolle, "Differentiated underwater sensor network deployment," in *Proceedings of the OCEANS—Europe*, pp. 1–6, June 2007.

[27] Y. Wang, Y. Liu, and Z. Guo, "Three-dimensional ocean sensor networks: a survey," *Journal of Ocean University of China*, vol. 11, no. 4, pp. 436–450, 2012.

[28] P. V. Teixeira, D. V. Dimarogonas, K. H. Johansson, and J. Sousa, "Event-based motion coordination of multiple underwater vehicles under disturbances," in *Proceedings of the IEEE Sydney OCEANS*, pp. 1–6, May 2010.

[29] J. Liu, X. Han, M. Al-Bzoor et al., "Prediction assisted dynamic surface gateway placement for mobile underwater networks," in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC '12)*, pp. 139–144, 2012.

[30] S. Ibrahim, J. Liu, M. Al-Bzoor, J. H. Cui, and R. Ammar, "Towards efficient dynamic surface gateway deployment for underwater network," *Ad Hoc Networks*, vol. 11, no. 8, pp. 2301–2312, 2013.

[31] W. Alsalih, H. Hassanein, and S. Akl, "Placement of multiple mobile data collectors in underwater acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 8, no. 8, pp. 1011–1022, 2008.

[32] W. Alsalih, H. Hassanein, and S. Akl, "Delay constrained placement of mobile data collectors in underwater acoustic sensor networks," in *Proceedings of the 33rd IEEE Conference on Local Computer Networks (LCN '08)*, pp. 91–97, October 2008.

[33] G. A. Hollinger, U. Mitra, and G. S. Sukhatme, "Autonomous data collection from underwater sensor networks using acoustic communication," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '11)*, pp. 3564–3570, September 2011.

[34] D. P. Williams, "AUV-enabled adaptive underwater surveying for optimal data collection," *Intelligent Service Robotics*, vol. 5, no. 1, pp. 33–54, 2012.

[35] S. Yoon and C. Qiao, "Cooperative search and survey using Autonomous Underwater Vehicles (AUVs)," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 3, pp. 364–379, 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 583753, 8 pages
http://dx.doi.org/10.1155/2013/583753

Research Article

Maximizing Network Lifetime of Directional Sensor Networks Considering Coverage Reliability

Joon-Min Gil,[1] Jong Hyuk Park,[2] and Young-Sik Jeong[3]

[1] *School of Information Technology Engineering, Catholic University of Daegu, Gyeongbuk 712-702, Republic of Korea*
[2] *Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul 139-743, Republic of Korea*
[3] *Department of Multimedia Engineering, Dongguk University, Seoul 100-715, Republic of Korea*

Correspondence should be addressed to Young-Sik Jeong; ysjeong@dongguk.edu

Received 8 July 2013; Accepted 17 November 2013

Academic Editor: Sana Ullah

Directional sensor networks composed of a large number of directional sensors equipped with a limited battery and with a limited angle of sensing have recently attracted attention. Maximizing network lifetime is a challenge in developing energy-efficient directional sensor networks, while covering all of the targets in a given area. However, the existing schemes have considered target coverage problem as only the maximization of network lifetime. In real sensor networks, the quality of target coverage will be varied according to the detection probability of the sensors covering targets. In this paper, we address the *directional cover-sets with coverage reliability (DCCR) problem* of organizing directional sensors into a group of nondisjoint subsets to extend network lifetime maximally while maintaining networks' satisfied coverage reliability. For the DCCR problem, we first present a coverage reliability model that mainly takes into account the detection probability of each sensor in cover-sets and eventually supports coverage reliability for target coverage. We also develop a heuristic algorithm called directional coverage and reliability (DCR) greedy algorithm to solve the DCCR problem. To verify and evaluate the algorithm, we conduct simulations and show that it extends network lifetime to a reasonable degree while guaranteeing the minimum coverage reliability.

1. Introduction

Recently, sensor networks have attracted considerable research interest due to their vast and significant applications, such as physical phenomenon or target detection, classification, and tracking [1–4]. Directional sensor networks (such as radar or image/video sensor networks [5, 6]) have different features than conventional sensor networks in which omnidirectional sensors are used. A directional sensor has directional coverage that can sense only in the direction of its orientation. These sensors have a limited sensing angle due to constraint related to manufacturing, size, and cost [7, 8]. Each directional sensor can sense only a sector of the disk, centered on itself, with the radius being equal to the sensing range.

The target coverage in directional sensor networks is determined by both the location and orientation of the sensors. This makes the network more complex. In addition, a power saving becomes a more critical issue because each sensor has only a limited amount of power and in most cases it is difficult or impossible to replace or recharge the batteries [9]. This issue is commonly resolved by using a sensor wake-up scheduling scheme by which some sensors remain active state to provide sensing services, while the others sleep state to conserve energy. Intuitively, if certain sensors share common sensing regions, some sensors can be switched into sleep state to conserve energy.

However, the most existing scheduling schemes have considered target coverage problem as only the maximization of network lifetime. In real sensor networks, the quality of target coverage will be varied according to the detection probability of the sensors covering targets. This is because each sensor has the different sensing capability by the signal attenuation rate of targets; that is, as the distance between a sensor and its interesting target increases, the signal attenuation increases. Therefore, coverage reliability is another important factor in

target coverage problem, which can improve the quality of services of directional sensor networks. In this paper, we consider a sensor scheduling problem to maximize network lifetime while maintaining the target coverage satisfying a specific requirement of coverage reliability.

We propose a new problem, called the directional cover-sets with coverage reliability (DCCR) problem, the objective of which is to maximize the lifetime of a directional sensor network while continuously monitoring all targets in a specific level of coverage reliability. Our strategy to solve this problem is to group all of the deployed directional sensors into a number of subsets, each of which should cover all of the targets. Only one subset is active at any given time; all others go into a sleep state. We call this type of network organization directional cover-sets with coverage reliability. The cover-sets need not be disjoint and each of them can be activated successively one by one. Due to the NP-completeness of the DCCR problem we designed a heuristic algorithm to solve the problem, called directional coverage and reliability (DCR) greedy algorithm, which uses a greedy method to generate the maximum number of cover-sets satisfying coverage reliability. (In this paper, we omit the complete proof of the fact that the DCCR problem is NP-complete.) To verify and evaluate the proposed algorithm, we conducted simulations, which showed that the proposed algorithm extends network lifetime to a reasonable degree while guaranteeing a minimum reliability.

The remainder of this paper is organized as follows. Section 2 presents a brief review of related work. In Section 3, we introduce our directional sensor and network model. In Section 4, we present coverage reliability model based on detection probability and describe the DCCR problem. Section 5 describes the proposed greedy algorithm to solve the DCCR problem. In Section 6, we discuss the performance evaluation of the proposed algorithm with simulations. Finally, Section 7 concludes the paper.

2. Related Work

For omnidirectional sensor networks, many scheduling algorithms for prolonging network lifetime while guaranteeing target coverage have been studied [10–12]. The goal of target coverage is to make each target in the physical space of interest locate within the sensing range of at least one sensor. Cardei and Du [13] introduced the target coverage problem, where disjoint sensor sets are modeled as disjoint cover sets, such that every cover set completely monitors all of the targets. This was called the *maximum set covers (MSC) problem* and was shown to be NP-complete. Requirements related to this problem were alleviated in [14], where sensors were not restricted to participation only in disjoint sets; that is, a sensor could be active in more than one set.

The scheduling problems in directional sensor networks have recently attracted a great deal of interest. Compared to omnidirectional sensors, directional sensors are obviously different in that their coverage region is determined by both location and orientation. Therefore, the target coverage problem aiming at directional sensors will be more complicated than that focusing on omnidirectional sensors [15]. Previous studies regarding the coverage of directional sensor networks also aimed to maximize network lifetime by finding cover sets to cover all targets [7, 8, 16]. The initial work relevant to the coverage issue in directional sensor networks was presented in [7]. The authors formulated the *maximum coverage with minimum sensors (MCMS) problem*, in which coverage in terms of the number of targets to be covered is maximized, while the number of sensors to be activated is minimized. The genetic algorithms, based on evolutionary search techniques, were applied in [8, 15] to solve target coverage problem in directional sensor networks.

From different points of view, He et al. [17] aimed to deploying sensors to cover-sets including network reliability in target coverage problem. To achieve this aim, authors introduced the failure probability associated with each omnidirectional sensor into sensor networks and tried to improve the network reliability by using the similar solution used to solve MSC problem. In contrast to that work, our work focuses on the maximization of network lifetime and the improvement of network reliability in directional sensor networks. Moreover, based on the detection probability of each directional sensor, the coverage reliability that means how reliable directional sensors in cover-sets can monitor all targets with a specific level is significantly considered.

3. Directional Sensor and Network Model

3.1. Directional Sensor Model. Here, we describe directional sensor model used in this paper. A sensor has W orientations and operates in only one orientation at any time; the active sensing region is determined by the chosen orientation. We do not make any assumptions regarding the shape of sensing regions, except that each sensing region is constrained by an angle of view and it is also closed, connected, and without holes.

A directional sensor s_i is represented by the six-tuple $\langle P_i, R_s, o_{i,j}, E_0(s_i), \omega\rangle$, where P_i is the location of the sensor s_i, R_s is the sensing range, $o_{i,j}$ is jth orientation of sensor s_i ($j = 1, 2, \ldots, W$, i.e., the number of orientations operated by a directional sensor is W), $E_0(s_i)$ is the initial energy of the sensor s_i, and ω is the angle of view. In this paper, we assume that all deployed sensors are homogeneous in terms of sensing range, angle of view, and number of orientations. We also assume that each sensor is aware of its location by using an arbitrary localization method [18–20] and a directional sensor network is connected due to the large communication range of sensors without considering the connectivity issue of sensors.

3.2. Network Model. Let us consider a directional sensor network composed of N sensors. All sensors are randomly scattered to cover M targets with fixed locations in a two-dimensional plane. We define $S = \{s_1, s_2, \ldots, s_N\}$ as the set of N sensors and $T = \{t_1, t_2, \ldots, t_M\}$ as the set of M targets. Let O be the set of all $o_{i,j}$ for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, W$ and $O(t_m)$ be the set of orientations that cover a target t_m ($m = 1, 2, \ldots, M$). A sensor is *active* if it is selected

to cover at least one target. A sensor that is not active goes into the *sleep* state. In this paper, directional sensor activity scheduling refers to determining the state of the deployed directional sensors (active or sleep).

4. Directional Cover-Set with Coverage Reliability Problem

In this section, we present the directional cover-sets with coverage reliability (DCCT) problem based on the directional sensor and network model described in the previous section.

4.1. Detection Probability Model. In real environment, each directional sensor has the detection probability of targets [21]. To take coverage reliability into consideration, we first describe the detection probability of targets in directional sensor networks.

To model the detection probability of targets, we modify the detection probability defined in [21] to accommodate directional sensors. We assume that the sensing range of each directional sensor s_i is a form of disk, centered at s_i, with radius R_s and s_i makes its orientation detection of whether a target t_m is covered. The orientation detection function $f(\phi(s_i, t_m))$ is given by

$$f\left(\phi\left(s_i, t_m\right)\right) = \begin{cases} 1, & \text{if } \phi_{s_i} \le \phi\left(s_i, t_m\right) \le \phi_{s_i} + \omega, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $\phi(s_i, t_m)$ is an angle between s_i and t_m, ϕ_{s_i} is an orientation angle of s_i, and ω is the angle of view.

We also assume that each sensor s_i has a detection probability of a target t_m satisfying its orientation detection. The detection probability $\widehat{p}_{i,m}$ can be calculated by

$$\widehat{p}_{i,m} = p_{i,m} \cdot f\left(\phi\left(s_i, t_m\right)\right), \tag{2}$$

where $p_{i,m}$ is a detection probability that depends on the signal propagation of the target t_m without any consideration of satisfying the orientation detection of the sensor s_i. Using a general signal propagation model where the signal parameter θ attenuates along with the signal propagation, we can calculate $p_{i,m}$ as follows:

$$p_{i,m} = \Pr\left[\frac{\theta}{d_i} + n_i \ge A\right] = Q\left(\frac{A - \theta/d_i}{\sigma_i}\right), \tag{3}$$

where d_i is the Euclidean distance between s_i and t_m, and A is a threshold determining the minimum signal strength that can be correctly decoded at s_i. n_i is the measurement noise and is assumed that it follows a Gaussian distribution with zero mean and variance σ_i^2. $Q(\cdot)$ is the Q-function defined by

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \tag{4}$$

If all directional sensors are homogenous physically, the noise variance σ_i^2 of each sensor s_i can be assumed to be identical. Then, we can rewrite (2) as follows:

$$\widehat{p}_{i,m} = Q\left(\frac{A - \theta/d_i}{\sigma}\right) \cdot f\left(\phi\left(s_i, t_m\right)\right). \tag{5}$$

From (5), we can say that the directional sensor s_i covers the target t_m if the received signal of s_i is larger than the threshold A and an angle between s_i and t_m is within a sector.

4.2. Coverage Reliability. As we described in the previous section, each sensor s_i has a detection probability when it covers a target t_m. The detection probability means how well s_i can detect the data emitted by t_m and is associated with the signal propagation of t_m. Using the detection probability $\widehat{p}_{i,m}$ defined by (5), we can obtain the coverage reliability of a target t_m covered by more than at least one directional sensor (i.e., by $O(t_m)$). There are many fusion techniques that can be used to derive the coverage reliability from the detection probability of each directional sensor. In this paper, the coverage reliability of a target t_m, $\widehat{P}_m$, is given by

$$\widehat{P}_m = 1 - \prod_{o_{i,j} \in O(t_m)} \left(1 - \widehat{p}_{i,m}\right), \tag{6}$$

where $\widehat{p}_{i,m}$ is a detection probability that a sensor s_i having an orientation $o_{i,j}$ covers t_m.

Now, we introduce the coverage reliability of a given cover-set that all targets can be reliably monitored. Using the coverage reliability of a target defined by (6), we can obtain the coverage reliability of a cover-set $\Upsilon(\tau_k)$, where τ_k is a given lifetime for the cover-set. The coverage reliability of the cover set $\Upsilon(\tau_k)$, $R(\tau_k)$, is given by

$$R\left(\tau_k\right) = \arg\min\left\{\widehat{P}_1, \widehat{P}_2, \ldots, \widehat{P}_M\right\}, \tag{7}$$

where M represents the number of targets.

Using the coverage reliability $R(\tau_k)$ obtained by (7), we can guarantee that all targets are reliably monitored with the coverage reliability of more than $R(\tau_k)$ by orientations in $\Upsilon(\tau_k)$.

4.3. DCCR Problem Formulation. In this section, we present the directional cover-sets with coverage reliability (DCCR) problem, which is a classical optimization problem.

As an orientation $o_{i,j}$ can belong to multiple cover-sets until the lifetime of the sensor s_i completely expires, we can define a Boolean variable $b_{i,j,k}$ as follows:

$$b_{i,j,k} = \begin{cases} 1, & \text{if } o_{i,j} \in \Upsilon\left(\tau_k\right), \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

By using (8), we define the DCCR problem as follows: for given a set of targets T with fixed locations and the initial energy $E_0(s)$ of each sensor in a set of directional sensors S, find a set of cover-sets $\Upsilon(\tau_1), \Upsilon(\tau_2), \ldots, \Upsilon(\tau_X)$ with coverage reliabilities $R(\tau_1), R(\tau_2), \ldots, R(\tau_X)$ such that (1) the network lifetime, denoted as $L(S, T, \kappa)$, is maximized and (2) each coverage reliability is not smaller than a minimum coverage reliability κ. Mathematically, the DCCR problem is defined as

Maximize

$$L(S, T, \kappa) \equiv \sum_{k=1}^{X} \tau_k \tag{9}$$

subject to

$$\sum_{k=1}^{X} E\left(s_i, \Upsilon\left(\tau_k\right)\right) \leq E_0\left(s_i\right), \quad \forall s_i \in S, \tag{10}$$

$$\sum_{j=1}^{W} b_{i,j,k} \leq 1, \quad \forall s_i \in S,\ k = 1, \ldots, X, \tag{11}$$

$$\sum_{o_{i,j} \in O(t_m)} b_{i,j,k} \geq 1, \quad \forall t_m \in T,\ k = 1, \ldots, X, \tag{12}$$

$$R\left(\tau_k\right) \geq \kappa, \quad k = 1, \ldots, X, \tag{13}$$

where

$$b_{i,j,k} = \{0, 1\}, \qquad \tau_k \geq 0. \tag{14}$$

Equation (10) guarantees that the total energy consumed by each directional sensor s_i across all cover-sets is not larger than its initial energy. Equation (11) guarantees that one directional sensor in a cover-set operates in at most one orientation. Equation (12) guarantees that each target is covered by at least one orientation in a cover-set. Finally, Equation (13) guarantees that all targets are reliably monitored by successive cover-sets in more than a certain level. Given a directional sensor network deployed in an area, the number of cover-sets, denoted by X, is finite but unknown. It should be noted that a sensor can appear in different cover-sets; that is, the sets of directional sensors in different cover-sets need not be disjoint. In the following section, we present a greedy algorithm to find such the cover-sets maximally.

5. The Proposed Greedy Algorithm to Solve the DCCR Problem

In this section, we propose a new heuristic algorithm, DCR greedy, to solve the DCCR problem. DCR greedy algorithm uses a greedy method to produce cover-sets and their coverage reliabilities by finding the cover-sets with coverage reliability in a given deployment of sensors and targets, based on the DCCR problem presented in Section 4.3.

The proposed algorithm is similar to those proposed previously [14, 17] but modified to capture the characteristics of directional sensor networks. Our algorithm takes as the input parameters S, T, $E_0(s_i)$ of each sensor s_i, the minimum coverage reliability κ, and the angle of view ω. Each cover-set operates for a fixed amount of time τ, unless some sensors in the cover-set die due to a lack of power. The output of the proposed algorithm is a sequence of cover-sets $\Upsilon(\tau_1), \Upsilon(\tau_2), \ldots, \Upsilon(\tau_X)$ and their coverage reliabilities $R(\tau_1), R(\tau_2), \ldots, R(\tau_X)$. It is noted that X should be maximized. By using the output, we can obtain the network lifetime of a given network of directional sensors, each of which faces an arbitrary orientation at the initial time of deployment as follows: $\tau_1 + \tau_2 + \cdots + \tau_X$. When an identical value for all τ_k $(k = 1, \ldots, X)$ is used as τ, the output of the proposed algorithm becomes $X \cdot \tau$. Moreover, the output of the proposed algorithm guarantees that a given network of directional sensors can reliably operate for $X \cdot \tau$ while keeping a minimum coverage reliability.

The pseudocode of the algorithm is shown in Algorithm 1. The following notations are used:

(i) S_l: set of live sensors;

(ii) O_l: orientations of live sensors;

(iii) $T(o_{i,j})$: set of targets covered by an orientation $o_{i,j}$;

(iv) x: index of cover-sets;

(v) $E_r(s)$: residual energy of a sensor s;

(vi) T_c: coverage reliability of a critical target t_c.

The algorithm repeatedly builds cover-sets and their coverage reliability until a cover-set is not found any more due to energy exhaustion of sensors. The process to find one cover-set and its coverage reliability stops once the entirety of each target is covered by at least one orientation of live sensors. The algorithm consists of the following steps.

Step 1. Initialize the variables S_l, O_l, and x used in the algorithm and the residual energy of each sensor (lines 1–7).

Step 2. To construct a cover-set, increase x by 1 and initialize the relevant variables such as *TARGETS*, τ_x, and $\Upsilon(\tau_x)$ (line 9).

Step 3. A *critical target* t_c is selected and defined as the most sparsely covered target, both in terms of number of sensors as well as the residual energy of those sensors [14]. After the critical target is selected, (6) is used to calculate the T_c that denotes a coverage reliability of the critical target (line 11). The critical target is a bottleneck in the viewpoint of network lifetime; that is, when the energy of the sensors covering the critical target is completely exhausted, the target cannot be covered and hence the network lifetime will be terminated.

Step 4. Initialize the variable O_c and insert all orientations covering the critical target t_c into O_c (lines 12-13).

Step 5. Determine whether the critical target t_c is reliable or not (line 14). If T_c is less than a minimum coverage reliability κ (i.e., the critical target is not as reliable as κ), proceed the next step to find another orientation covering the critical target except already selected orientations; otherwise go to Step 8.

Step 6. Select the orientation $o_{s,t}$ with the maximum profit value $W(o_{i,j})$ among the orientations covering the critical target t_c (line 15). Various profit functions can be defined and we consider two kinds of criteria; one is a reliability profit

Input parameters: $S, T, E_0(s_i), \kappa, \omega, \tau$
(1) $S_l = S, O_l = \emptyset, x = 0$
(2) **for** each $s_i \in S_l$ **do**
(3) $E_r(s_i) = E_0(s_i)$
(4) **for** each orientation $o_{i,j}$ operated by s_i **do**
(5) $O_l = O_l \cup \{o_{i,j}\}$
(6) **end for**
(7) **end for**
(8) **while** $O_l \neq \emptyset$ **and** $\bigcup_{o_{i,j} \in O_l} T(o_{i,j}) = T$ **do**
(9) $x = x + 1, TARGETS = T, \tau_x = \tau, \Upsilon(\tau_x) = \emptyset$
(10) **while** $TARGETS \neq \emptyset$ **do**
(11) Find a critical target $t_c \in TARGETS$ and calculate T_c.
(12) $O_c = \emptyset$
(13) Find all orientations $\in O_l$ that cover t_c and insert them into O_c
(14) **while** $T_c < \kappa$ **do**
(15) Select $o^*_{s,t} \in O_c$ with the maximum profit $W(o^*_{s,t})$
(16) $\Upsilon(\tau_x) = \Upsilon(\tau_x) \cup \{o^*_{s,t}\}$
(17) **for** each orientation $o_{i,j} \in O_l$ **do**
(18) **if** $i = s$ **then**
(19) $O_l = O_l - \{o_{i,j}\}$
(20) **end if**
(21) **end for**
(22) Recalculate T_c
(23) **end while**
(24) **for** each $t_m \in TARGETS$ **do**
(25) **if** t_m is covered by $o_{i,j} \in \Upsilon(\tau_x)$ **and** $T_m \geq \kappa$ **then**
(26) $TARGETS = TARGETS - \{t_m\}$
(27) **end if**
(28) **end for**
(29) **end while**
(30) **for** each $s \in \Upsilon(\tau_x)$ **do**
(31) $E_r(s) = E_r(s) - E(s, \Upsilon(\tau_x))$
(32) **if** $E_r(s) \leq 0$ **then**
(33) $S_l = S_l - \{s\}$
(34) **end if**
(35) **end for**
(36) $O_l = \bigcup_{j=1}^{W} \{o_{i,j}\}$ **for each** sensor s_i in S_l
(37) **end while**
(38) **return** $\Upsilon(\tau_1), \Upsilon(\tau_2), \ldots, \Upsilon(\tau_X)$ and $R(\tau_1), R(\tau_2), \ldots, R(\tau_X)$

Algorithm 1: DCR-greedy algorithm.

and the other is an energy profit. In this paper, we use the following function:

$$f(o_{i,j}, t_c) = \alpha \cdot f_1(o_{i,j}, t_c) + (1 - \alpha) \cdot f_2(o_{i,j}, t_c), \quad (15)$$

$$f_1(o_{i,j}, t_c) = \prod_k (1 - p_{i,k}) \cdot p_{i,c}, \quad (16)$$

$$f_2(o_{i,j}, t_c) = E_r(s_i), \quad (17)$$

$$W(o_{i,j}) \equiv \arg\max_{o_{i,j}} f(o_{i,j}, t_c), \quad (18)$$

where $0 < \alpha < 1$. In (15), $f_1(o_{i,j}, t_c)$ denotes a reliability profit function to get the coverage reliability of the critical target t_c that is calculated by the detection probability of the orientation $o_{i,j}$ covering the critical target, and $f_2(o_{i,j}, t_c)$ denotes an energy profit function to get the residual energy of the orientation $o_{i,j}$ covering the critical target. In (16), $p_{i,k}$ represents the detection probability of the sensor s_i whose the orientation $o_{i,j}$ covers another target t_k while covering the critical target t_c and $p_{i,c}$ represents the detection probability of the sensor s_i whose orientation $o_{i,j}$ covers the critical target t_c. In terms of the reliability profit, the orientation $o_{i,j}$ is selected such that it has a maximum detection probability for the critical target and at the same time minimizes detection probability for other targets except the critical target. By choosing a proper value of α, the orientation will be selected such that it has higher contribution than any other orientations for detecting the critical target and the sensor s_i with the selected orientation has more residual energy available.

Step 7. Once the orientation $o_{s,t}$ is selected, it is added to the current cover-set $\Upsilon(\tau_x)$ (line 16) and all orientations of the

sensor s_s operating the selected orientation $o_{s,t}$ are removed from the live orientation set O_l (lines 17–21).

Step 8. Recalculate the coverage reliability T_c (line 22) and go to Step 5.

Step 9. All of the targets that are covered by $o_{s,t}$ and at the same time are not less than κ, are removed from the current target set, *TARGETS* (lines 24–28).

Step 10. After the cover-set $\Upsilon(\tau_x)$ is built, the residual energy of each sensor in $\Upsilon(\tau_x)$ is updated (line 31).

Step 11. Dead sensors are removed from the set of live sensors (lines 32–34). If a sensor has no residual energy, it becomes a dead sensor.

Step 12. Before going to line 8 and repeating the above steps from Step 2, the set of available orientations O_l is updated based on live sensors (line 36).

6. Performance Evaluation

In this section, we evaluate and analyze the performance of the proposed DCR greedy algorithm through simulations.

6.1. Simulation Environment. We simulated a stationary network with sensors and targets randomly deployed in an area of 500 m × 500 m. Our simulation environment assumes that the different numbers of targets (M = 5 and 15) are uniformly deployed in the area and the different numbers of sensors (N = 20, 50, 80, and 110) are randomly scattered to cover the targets. The initial energy of each directional sensor ($E_0(s_i)$) was set to 1.0 and the active time of all cover-sets (τ) is set to 0.1. The values of three parameters for detection probability were chosen as follows: A = 2.0, θ = 250, and σ = 1.65. We assume that all directional sensors can sense one of three orientations (W = 3) with a sensing range of 250 m and there is no overlap with the other two orientations; that is, the angle of view (ω) is fixed to $2\pi/3$. The minimum coverage reliability (κ) ranged from 0.3 to 0.9 is used to study the effects on network lifetime.

The proposed DCR greedy algorithm is evaluated with the following two effects.

(i) Minimum coverage reliability: this is used to investigate whether our DCRgreedy algorithm can solve the DCCR problem defined in this paper. We then analyze the performance of our algorithm for the different values of minimum coverage reliability in terms of network lifetime.

(ii) Orientation with maximum profit: as the value of α in the maximum profit function f defined in Section 5, we use one of the following three ones, each of which represents the different strategy to select an orientation with maximum profit.

(a) α = 0.0: consider only the coverage reliability of targets.

(b) α = 0.5: consider both the coverage reliability of targets and the residual energy of directional sensors at the same time.

(c) α = 1.0: consider only the residual energy of directional sensors.

In the next subsections, we present the simulation results to analyze the effect of these factors on the network lifetime. The results presented here have been averaged over 30 simulation runs based on randomly generated networks.

6.2. Effect of Minimum Coverage Reliability. We investigated the impact of minimum coverage reliability on network lifetime. For this investigation, we fixed the value of α to 0.5, and network lifetime was evaluated for four values of minimum coverage reliability (i.e., κ = 0.3, 0.5, 0.7, and 0.9).

Figure 1 shows the comparison of network lifetime with varying minimum coverage reliability. The number of directional sensors (N) varied between 20 and 110 and the number of targets (M) was fixed as 5 (Figure 1(a)) and 15 (Figure 1(b)). The network lifetime almost linearly increased with the number of directional sensors regardless of the number of targets, because a larger number of directional sensors can be scheduled to cover the given targets. It should be noted that an increase in the number of directional sensors leads to more cover-sets, as more directional sensors are available for target coverage.

On comparing network lifetime for four values of minimum coverage reliability (κ), we can see that when high minimum coverage reliability is used, the resulting network lifetime tends to be low. This is natural because the cover-sets with high minimum coverage reliability are organized with more orientations than those with low one. Therefore, our DCR greedy algorithm can find the cover-sets satisfying a minimum coverage reliability when greedily selecting an energy-efficient orientation of a directional sensor to organize a cover-set.

6.3. Effect of Orientation with Maximum Profit. We investigated the effect of the orientation with maximum profit on network lifetime. The network lifetime is compared in accordance with the number of directional sensors for different orientation selection strategies. The number of directional sensors ranged from 20 to 110 was used to cover 5 and 15 targets, respectively. For simplicity, we set the minimum coverage reliability to 0.7. Three kinds of α to select the orientation with maximum profit were used in this evaluation.

Figure 2 shows the effect of the orientation with maximum profit on network lifetime. It is shown that the network lifetime is longest when α is 0.5, regardless of the number of directional sensors and targets. This indicates that our DCR greedy algorithm can maximally extended the network lifetime when it considers both the coverage reliability of targets and the residual energy of directional sensors at the same time.

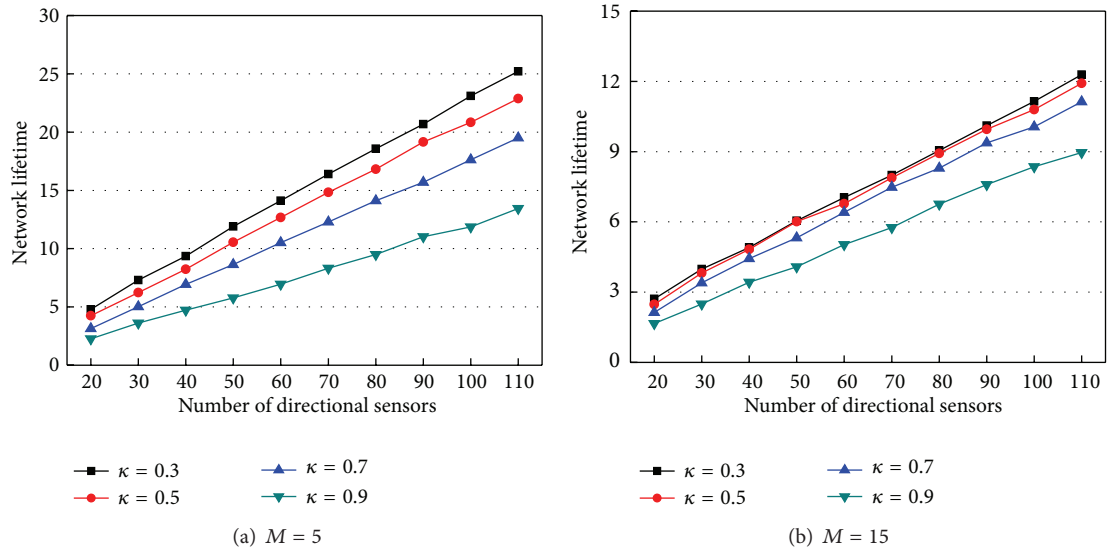


Figure 1: Comparison of network lifetime with varying minimum coverage reliability.

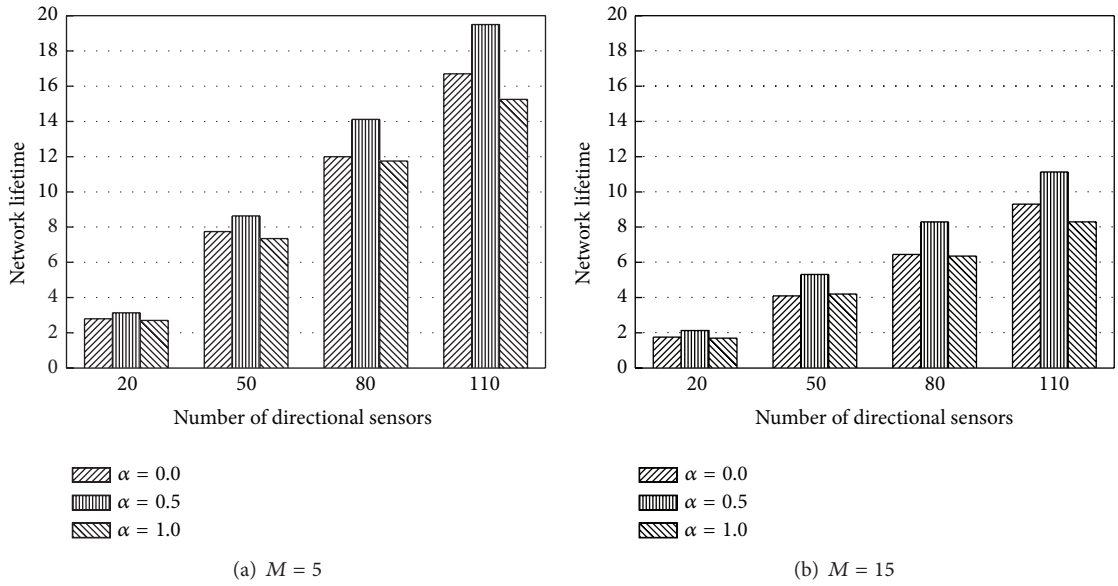


Figure 2: Comparison of network lifetime according to different orientation selection strategies.

7. Conclusions

We have proposed a new scheduling problem called the DCCR problem and developed a heuristic algorithm called DCR greedy algorithm to schedule active directional sensors to maximize network lifetime. The active sensors in cover-sets scheduled by the algorithm can cover all of the targets while maintaining networks' satisfied coverage reliability. The algorithm can extend the lifetime of a directional sensor network maximally while continuously and reliably monitoring all targets with a specific level of coverage reliability. Throughout simulation studies, two effects regarding minimum coverage reliability and orientation with maximum profit were used to investigate the performance of our algorithm. The simulation results showed that the DCR greedy algorithm is suitable for solving the DCCR problem regardless of the minimum coverage reliability used. The DCR greedy algorithm achieves high performance in terms of network lifetime when it considers both the coverage reliability of cover-sets and the residual energy of directional sensors.

Acknowledgments

This research is supported by the MSIP (Ministry of Science, ICT and Future Planning), Republic of Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-4007) supervised by

the NIPA (National IT Industry Promotion Agency) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012000534).

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] M. Huadong and L. Yonghe, "Correlation based video processing in video sensor networks," in *Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing*, pp. 987–992, June 2005.

[3] C. Huang, R. H. Cheng, S. R. Chen, and C. I. Li, "Enhancing network availability by tolerance control in multi-sink wireless sensor networks," *Journal of Convergence*, vol. 1, no. 1, pp. 15–22, 2010.

[4] G. Zhao and A. Kumar, "Lifetime-aware geographic routing under a realistic link layer model in wireless sensor network," *Journal of Information Technology, Communications and Convergence*, vol. 1, no. 3, pp. 297–317, 2011.

[5] H. Ma and Y. Liu, "On coverage problems of directional sensor networks," in *Proceedings of the 1st International Conference on Mobile Ad-hoc and Sensor Networks*, pp. 721–731, 2005.

[6] M. Rahimi, R. Baer, O. Iroezi et al., "Cyclops: in situ image sensing and interpretation in wireless sensor networks," in *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*, pp. 192–204, 2005.

[7] J. Ai and A. A. Abouzeid, "Coverage by directional sensors in randomly deployed wireless sensor networks," *Journal of Combinatorial Optimization*, vol. 11, no. 1, pp. 21–41, 2006.

[8] J. Wang, C. Niu, and R. Shen, "Priority-based target coverage in directional sensor networks using a genetic algorithm," *Computers and Mathematics with Applications*, vol. 57, no. 11-12, pp. 1915–1922, 2009.

[9] M. Ilyas and I. Mahgoub, *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, Florida Atlantic University, 2004.

[10] M. Cardei and J. Wu, "Energy-efficient coverage problems in wireless ad-hoc sensor networks," *Computer Communications*, vol. 29, no. 4, pp. 413–420, 2006.

[11] J. Chen and X. Koutsoukos, "Survey on coverage problems in wireless ad hoc sensor networks," in *Proceedings of the IEEE SouthEast Conference*, pp. 22–25, Richmond, Va, USA, March 2007.

[12] C.-F. Huang and Y.-C. Tseng, "A survey of solutions to the coverage problems in wireless sensor networks," *Journal of Internet Technology*, vol. 6, no. 1, pp. 1–8, 2005.

[13] M. Cardei and D.-Z. Du, "Improving wireless sensor network lifetime through power aware organization," *Wireless Networks*, vol. 11, no. 3, pp. 333–340, 2005.

[14] M. Cardei, M. T. Thai, Y. Li, and W. Wu, "Energy-efficient target coverage in wireless sensor networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '05)*, pp. 1976–1984, Miami, Fla, USA, March 2005.

[15] J.-M. Gil and Y.-H. Han, "A target coverage scheduling scheme based on genetic algorithms in directional sensor networks," *Sensors*, vol. 11, no. 2, pp. 1888–1906, 2011.

[16] Y. Cai, W. Lou, M. Li, and X.-Y. Li, "Energy efficient target-oriented scheduling in directional sensor networks," *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1259–1274, 2009.

[17] J. He, N. Xiong, Y. Xiao, and Y. Pan, "A reliable energy efficient algorithm for target coverage in wireless sensor networks," in *Proceedings of the 30th International Conference on Distributed Computing Systems Workshops (ICDCSW '10)*, pp. 180–188, June 2010.

[18] J. Liu and S. H. Chung, "An efficient load balancing scheme for multi-gateways in wireless Mesh networks," *Journal of Information Processing Systems*, vol. 9, no. 3, pp. 365–378, 2013.

[19] Y. Liu, Z. Yang, X. Wang, and L. Jian, "Location, localization, and localizability," *Journal of Computer Science and Technology*, vol. 25, no. 2, pp. 274–297, 2010.

[20] D. Kumar, T. C. Aseri, and R. B. Patel, "Multi-hop communication routing (MCR) protocol for heterogeneous wireless sensor networks," *International Journal of Information Technology, Communications and Convergence*, vol. 1, no. 2, pp. 130–145, 2011.

[21] B. Wang, *Coverage Control in Sensor Networks*, Springer, 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 829691, 6 pages
http://dx.doi.org/10.1155/2013/829691

Research Article

A QoS Model for a RFID Enabled Application with Next-Generation Sensors for Manufacturing Systems

Anna Kang,[1] Jong Hyuk Park,[2] Leonard Barolli,[3] and Hwa-Young Jeong[4]

[1] *Department of Multimedia Engineering, Dongguk University, Seoul 100-715, Republic of Korea*
[2] *Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul 139-743, Republic of Korea*
[3] *Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), Fukuoka 811-0295, Japan*
[4] *Humanitas College, Kyung Hee University, Seoul 130-701, Republic of Korea*

Correspondence should be addressed to Hwa-Young Jeong; jhymichael@gmail.com

Received 15 September 2013; Accepted 18 October 2013

Academic Editor: Sana Ullah

RFID is a new trend in industrial technology such as manufacturing system and product flow management in the supply chain. A successful RFID application can enhance the industrial technology change in organizations and help to manage growth in an increasingly competitive environment. Recently, the RFID sensor tag has also recently received a great deal of attention. In general, the RFID sensor was designed to collect environmental information such as temperature, pressure, humidity, inclination, and acceleration into ordinary RFID tags for internal information processing. In this environment, to apply the RIFID technique to manufacturing systems successfully and efficiently, identifying accurate quality attributes and making a quality model for a RFID enabled application system are necessary. However the research in this area is lacking. This paper focuses on identification of the quality attributes for the RFID application system and making the corresponding model. Each criterion for the quality model was extracted in accordance with ISO/IEC 9126 and the DeLone and McLean model. The proposed quality model also considers the relationship between their criteria.

1. Introduction

Radio frequency identification (RFID) is one of several kinds of automatic identification sensor-based technologies consisting of three key elements: RFID tags, RFID readers, and a back-end database server that has the ability to identify or scan information with increased speed and accuracy. In a basic RFID system, the information transmitted in the air between the tag and the reader could easily be subject to interception and eavesdropping due to the nature of its radio transmission [1]. The application of a radio frequency identification (RFID) sensor tag or smart active label, which is a typical sensor-enabled semipassive RFID tag, has recently received a great deal of attention. RFID sensor tags are, in general, the integration of various sensors designed to collect environmental information, including temperature, pressure, humidity, inclination, and acceleration, into ordinary RFID tags, which are originally used only to identify products, with a film battery supplying power for internal information processing [2]. In this environment, radio frequency identification (RFID) has become a critical issue in the fields of manufacturing and logistics. This rapid development of RFID has led lots of companies to take a hard look at what RFID can do for commercial purposes. Moreover, RFID is regarded as a promising technology for the management of supply chain processes since it improves the efficiency of forecasting customer demands, production planning, managing inventory, and retail operations. Recently, RFID has become important in mobile and wireless communication technologies [3, 4] and has influenced various other industries also. But its range of application is going to be extended far beyond these areas. There is tremendous potential for applying it even more widely, and increasing numbers of companies have already started up pilot schemes or successfully applied it to real-world environments. Consequently, RFID is to be among the most exciting and fastest-growing technologies in terms of

the scope of application in the next generation of business intelligence. However, RFID vendors are complaining that the business is not growing as fast as expected [5] and that the main system for the management and control of the processes of the RFID device signals sometimes cannot be trusted. Therefore, the method for identifying the QoS (Quality of Service) for RFID enabled applications is crucial for RFID manufacturers.

Quality can be defined as the possession by a product of the conditions that make it suitable to meet the expressed or potential needs of its users [6]. To determine the system quality, quality metric models have been studied by many researchers. This research selects six attributes with 27 subcriteria in ISO/IEC 9126-1, which is the revision of the 1991 version (ISO/IEC 9126, 1991). ISO/IEC 9126-1 defines the terms for the system quality characteristics and how these characteristics are decomposed into subcharacteristics. ISO/IEC 9126-1, however, does not describe how any of these subcharacteristics could be measured. To address this issue, three more parts are extended: ISO/IEC 9126-2, ISO/IEC 9126-3, and ISO/IEC 9126-4. ISO/IEC 9126-2 defines external metrics which measure the behaviors of the computer based system. ISO/IEC 9126-3 defines internal metrics which measure the software itself. ISO/IEC 9126-4 defines quality in terms of metrics which measure the effects of using the software in a specific context of use. However, a drawback of the existing international standards is that they provide very general quality models and guidelines but are very difficult to apply to specific domain [7].

DeLone and McLean [8] become aware of the complex reality that surrounds the identification and definition of the IS (Information System) success concept. They conducted a large number of studies on IS success and presented a comprehensive and integrative model [9, 10].

This research aims to identify the quality attributes and make a QoS model for RFID sensor technology enabled application system. For this process, each criterion of the quality attributes was extracted by ISO/IEC 9126 and we added characteristics to take into consideration RFID sensor technology in manufacturing systems. To make a quality model, the research used DeLone and McLean's model to consider their relation between criteria. The model will support the guidelines when the system engineer or developer wants to develop a RFID enabled application system.

The rest of this paper is organized as follows. Section 2 reviews the related works for RFID enabled application systems with manufacturing systems. Section 3 explains criteria which take into account their characteristics, the factors, criteria, and their correlation for QoS model in manufacturing systems with RFID technology. Section 4 describes the proposed QoS model by experiment. Finally, Section 5 presents the conclusions.

2. A RFID Enabled Application System

2.1. Manufacturing Systems and Their Processes. There is always a need for human expertise to combine components and system elements in order to build a purposeful manufacturing system that can produce a targeted class of products. In the content, a manufacturing system has to combine separate components or elements together to form a whole system and these synthesis activities are always related to human activities for creating artifacts [11]. Traditional mass manufacturing, also known as repetitive flow production, series production, or flow production, is the production of a large number of standardized products usually on automated production lines. The manufacturing procedure strictly follows a set of predefined standards that are generated from a test run of a small sample. These standards include those related to parts, human labor, processes, machinery operations, and the general working environment. Once the standards are defined and generated, mass produced goods are manufactured by strictly following these standards. The standards themselves are kept unchanged unless a large deviation in production occurs or a routinely scheduled test is arranged. Components that are part of the final product follow standards obtained during the testing phase and are treated uniformly at the mass production stage [12].

2.2. RFID and Sensor Technology. RFID is an automatic identification technology that can be used to provide a unique ID to a physical object. A typical RFID system consists of RFID reader(s), tags, RFID middleware, a RFID database, and RFID services as shown in Figure 1. Communication in RFID occurs through radio waves, where information from a tag to a reader or vice versa is sent via an antenna [13].

The benefits of integrating two complementary technologies—RFID and sensor—are substantial, as pointed out by many researchers. First of all, from the information acquisition perspective, the richer information available from the addition of "environment traceability" to "object traceability" facilitates better decision support and responsive localized management. If the RFID technology is embedded in a wireless sensor network (WSN), RFID tags and readers can build more intelligent networks by sharing the sensing, logic, and transmission capabilities of the sensor networks. For example, longer-distance transmission can be achieved with lower power consumption and less interference by utilizing multihop transmission and clustered reading capabilities provided by sensor network technologies [2]. One WSN may be composed of hundreds or thousands of miniature sensor nodes, or motes, which are fitted with an onboard processor. The low-cost battery-powered sensor nodes have an extremely limited energy supply, stringent processing and communications capabilities, and scarce memory. Sensor nodes are usually densely deployed in a sensor field in order to continuously monitor surrounding areas. In a sensor application, each sensor has the capability to collect data such as temperature, humidity, light condition, and so on, depending on targeted applications. After sensor nodes collect data, they can locally carry out some simple computations and collaboratively route data to a base station for analysis. A base station may be a fixed node or a mobile node capable of connecting WSNs to a communications infrastructure (e.g., the Internet) where users can have access to reported data. In order to reduce the amount of raw data transmitted to a base station and to save energy, sensor nodes often need to perform aggregation operations so that only processed

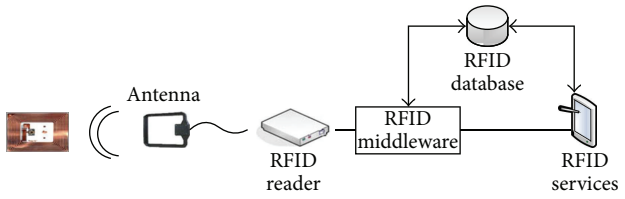


Figure 1: Typical RFID system.

information, for instance, the mean, max, or min of sensed raw data, is transmitted [14].

2.3. A RFID Enabled Application System. Basically RFID systems consist of tags and readers. Tags, also named transponders, are attached to the item being tracked and have data (e.g., an identification number or temperature) stored in their memory. Readers, also named interrogators, are the devices that read data from, and depending on the RFID system write data to tags and are connected to a network, like a local area network (LAN) to transport their data to, for example, a central database installed on a server [15]. Poon et al. [16] depicted the configuration of material test using RFID as shown in Figure 2.

In this environment, the reading performance of a RFID device is measured when the tags are placed on the front and back surfaces of various types of products in the actual environment.

In comparison to other well-known auto-ID technologies such as the barcode, RFID offers the following advantageous characteristics for the user [17].

(i) Unique identification: applying, for example, the "Electronic Product Code" (EPC) standards, RFID tags can identify classes of products as well as individual items.

(ii) No line of sight: RFID tags can be read without direct line of sight even if the tag is covered, dirty, or otherwise obscured from view.

(iii) Bulk reading: If they are in range of a reader, multiple RFID tags can be read at the same time.

(iv) Storage capacity: RFID tags can store significantly more information than just an identification number.

(v) Dynamic information: RFID tags with read-write capability allow information to be updated or changed whenever necessary.

3. A QoS Model for a RFID Enabled System

3.1. The System Quality with ISO/IEC Standards. RFID applications have evolved towards the need to provide information about the status of the item that its being uniquely identified. ISO/IEC 18000 defines a series of RFID air interface standards for item identification, operating at various frequencies and thus with different functionalities. The standard currently consists of seven parts. Part 1 is a reference architecture and definition of parameters to be standardized, while the other six parts describe air interfaces for various frequency

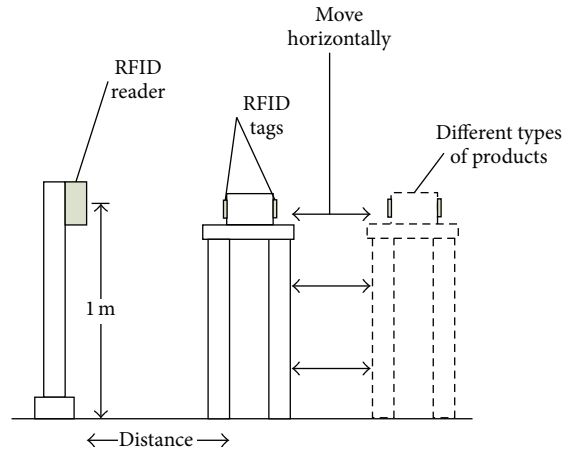


Figure 2: Configuration of material test.

ranges. The inclusion of sensors within the standard is mostly considered for parts 6 and 7, although most of the other parts make some provisions for the inclusion of sensor hardware. The objective of ISO/IEC 24753.2 is to provide common encoding rules for identifying sensors, their functions, and their delivered measurements (both simple and full-function sensors). It also defines the process rules for related functionality such as showing how to start and stop a particular sensor's monitoring function, how to access the sensor data, and how to carry out basic processing to convert the sensor data into meaningful information for an application. It specifies the physical interactions between interrogators and tags, the interrogator and tag operating procedures and commands, and the collision arbitration scheme used to identify a specific tag in a multiple-tag environment. ISO/IEC 18000-6 specifies four communication types as a simple sensor and full-function sensor [18].

(i) Simple sensor: a simple sensor is factory-programmed. Its objective is to provide a simple sensor data block (SSD) appended to the object-related unique identifier, using the delivery mechanism defined by the air protocol interface. The SSD includes information about the type of sensor (temperature, relative humidity, impact, tilt, and time-temperature integration are supported), configuration, and alarm status (on/off). A more complex sensor output than an on/off alarm status is possible, such as 8-bit sensor values.

(ii) Full-function sensor: Full-function sensors provide greater flexibility than simple sensors, supporting a greater variety of sensor types and measurement spans, enabling thresholds to be set within a wider range and enabling capturing and processing of different types of data. Unlike simple sensors, full-function sensors require a dedicated dialogue with the interrogator and may be programmed multiple times by the user.

The IEEE 1451 family of standards aims to provide a standard way of accessing any type of transducer regardless of

the type, manufacturer, and underlying information network. An IEEE 1451 smart transducer has the capabilities for self-identification, self-description, self-diagnosis, self-calibration, location awareness, time awareness, data processing, reasoning, data fusion, alert notification, standard based data formats, and communication protocols [18]. However, despite the many standards for RFID technique, there are a few standards for RFID system quality.

In an effort to identify attributes of system product quality that can be useful to developers, acquirers, and evaluators, a set of international standards have been developed. They are ISO/IEC 9126: Part 1 (quality model, 2001), Part 2 (external metrics, 2003), Part 3 (internal metrics, 2003), and Part 4 (quality in use metrics, 2004). ISO/IEC 9126-1 defines a quality model that includes six characteristics (functionality, reliability, usability, efficiency, maintainability, and portability), which are further subdivided into 27 subcharacteristics [19]. Table 1 shows ISO9126 quality model.

This research extracts the RFID system quality from ISO9126 and Lee et al.'s model [20], as shown in Table 2.

Particularly, the subcriteria were from Lee et al.'s model [20], such as read range, read accuracy, identification, and interference of functionality, data capacity of efficiency, and cost of business. Data capacity has 100s–1000s of characters. In read range, passive RFID case has Up to 25 feet and active RFID case has up to 100s of feet or more. The read rate has 10s, 100s or 1000s simultaneously. Read accuracy depends 90% on the relative orientations of the reader and tag antennas and their polarizations. In identification, it can uniquely identify each item/asset tagged in Interference, like the TSA (Transportation Security Administration) and some RFID frequency does not like metal and liquid. They can cause interference with certain RF frequencies. Cost normally has tag 5¢ RFID startup kit with RFID reader, antennas, alien gateway software, startup kit tag, and power supply/power cable for USD 2595.

4. QoS Model for a RFID System

We consolidated and synthesized the constructs using DeLone and McLean's model [21] as shown in Figure 3. Their model does not focus on the RFID system but proposes the relationships between each criterion for the general system. Their model consists of six distinct aspects of systems success: "system quality," "information quality," "use," "user satisfaction," "individual impact," and "organizational impact."

Through this model and the criteria in Table 2, the quality model for a RFID system is depicted as shown in Figure 4. This model consists of 6 criteria: functionality, reliability, usability, efficiency, maintainability, and business. Each criterion affects the relationship between user satisfaction and developer satisfaction.

Because some studies provide Kendall's correlation coefficients instead of Pearson's correlation coefficients, Kendall's correlation coefficients are transformed into Pearson's correlation coefficients by using the formula suggested by Kruskal [22, 23]:

$$\rho = \sin\left[\left(\frac{\pi}{2}\right)\tau\right]. \quad (1)$$

Table 1: ISO9126 quality model.

Criteria	Sub-criteria
Functionality	Suitability
	Accuracy
	Interoperability
	Compliance
	Security
Reliability	Maturity
	Recoverability
	Fault tolerance
Usability	Learnability
	Understandability
	Operability
Efficiency	Time behavior
	Resource behavior
Maintainability	Stability
	Analyzability
	Changeability
	Testability
Portability	Installability
	Conformance
	Replaceability
	Adaptability

Table 2: Proposed RFID system quality attributes.

Criteria	Sub-criteria
Functionality	Suitability
	Read accuracy
	Read range
	Read rate
	Identification
	Interface
Reliability	Maturity
	Recoverability
	Fault tolerance
Usability	Easy to use
	Understandability
	Operability
Efficiency	Time behavior
	Resource behavior
	Data capacity
Maintainability	Stability
	Analyzability
	Changeability
	Testability
Business	Cost

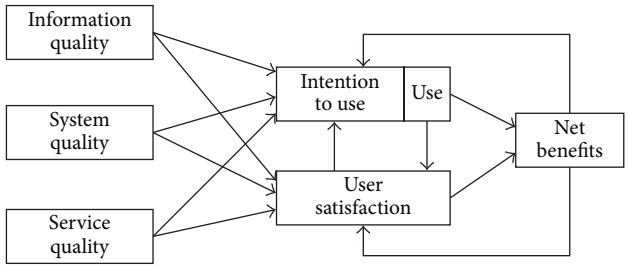


Figure 3: DeLone and McLean model [21].

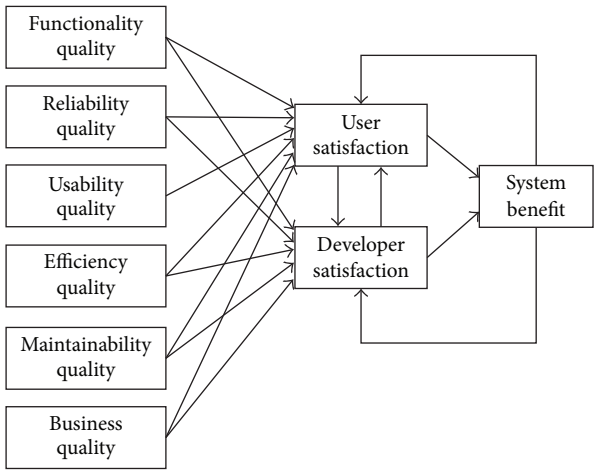


Figure 4: Quality model for a RFID system.

Table 3: Rank and Pearson's correlation coefficients for manufacturing system with RFID sensor technology.

Criteria	Rank	Sub-criteria	Pearson's correlation coefficients
Functionality	1	Interface	0.324
	2	Read rate	0.230
	3	Read range	0.212
	4	Suitability	0.189
	5	Read accuracy	0.125
	6	Identification	0.111
Reliability	1	Fault tolerance	0.284
	2	Recoverability	0.226
	3	Maturity	0.148
Usability	1	Operability	0.277
	2	Easy to use	0.204
	3	Understandability	0.185
Efficiency	1	Resource behavior	0.299
	2	Time behavior	0.222
	3	Data capacity	0.213
Maintainability	1	Stability	0.281
	2	Changeability	0.265
	3	Analyzability	0.207
	4	Testability	0.189
Business	1	Cost	0.273

Pearson's correlation coefficients of all criteria within each category are combined by using the following formula suggested by Hunter et al. [24] and Hwang et al. [23, 25]:

$$\rho = \frac{\sum(\rho n)}{\sum(n)}. \quad (2)$$

In order to evaluate each criterion's correlation, technical profiles of 80 participants of this study were distributed evenly between manufacturing system engineers (25.00%), system managers (25.00%), RFID sensor network engineers (25.00%), and experts of ISO/IEC standards (25.00%). Most of the participants had more than four years of experience in their positions (90.00%).

As for successful participants, user perceived benefits rank highly for two measures of success: Functionality (rank 1) and Efficiency (rank 1) as shown in Table 3. Actually, functionality has special characteristics such as interface for RFID sensor and suitability for RFID to manufacturing system.

5. Conclusions

RFID technology is a new trend for the industrial environment with IT (information technology). It has many benefits to us whether we are a user, a system developer, or an engineer. Therefore there are many standards for using the RFID sensor technique such as ISO/IEC 18000, ISO/IEC 24753.2, and IEEE 1451. However, they only consider system performance when the user or system engineer uses the RFID sensor technique for their system. In the contents, the research for RFID system quality model has been very lacking. This research focuses on making a quality model for a RFID enabled application system. Each criterion of the quality was borrowed from ISO9126 and Lee et al.'s model [20]. The quality model consists of 6 criteria; Functionality, reliability, usability, efficiency, maintainability, and business. All the criteria affect user satisfaction and developer satisfaction. And, finally, when the manufacturing system satisfies user and developer, RFID system benefit can be obtained.

Acknowledgment

This research is supported by the Ministry of Science, ICT and Future Planning (MSIP), Republic of Korea, under the Information Technology Research Center (ITRC) support program (NIPA-2013-H0301-13-4007) supervised by the National IT Industry Promotion Agency (NIPA).

References

[1] E.-J. Yoon, "Improvement of the securing RFID systems conforming to EPC class 1 generation 2 standard," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1589–1594, 2012.

[2] Y.-S. Kang, H. Jin, O. Ryou, and Y.-H. Lee, "A simulation approach for optimal design of RFID sensor tag-based cold chain systems," *Journal of Food Engineering*, vol. 113, pp. 1–10, 2012.

[3] J. H. Chong, C. K. Ng, N. K. Noordin, and B. M. Ali, "Dynamic transmit antenna shuffling scheme for MIMO wireless communication systems," *Journal of Convergence*, vol. 4, no. 1, pp. 7–14, 2013.

[4] A. Sinha and D. K. Lobiyal, "Performance evaluation of data aggregation for cluster-based wireless sensor network," *Human-Centric Computing and Information Sciences*, vol. 3, article 13, 2013.

[5] C.-S. Lin, L.-S. Chen, and C.-C. Hsu, "An innovative approach for RFID product functions development," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15523–15533, 2011.

[6] P. Papetti, C. Costa, F. Antonucci, S. Figorilli, S. Solaini, and P. Menesatti, "A RFID web-based infotracing system for the artisanal Italian cheese quality traceability," *Food Control*, vol. 27, no. 1, pp. 234–241, 2012.

[7] K. K. F. Yuen and H. C. W. Lau, "A fuzzy group analytical hierarchy process approach for software quality assurance management: fuzzy logarithmic least squares method," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10292–10302, 2011.

[8] W. H. DeLone and E. R. McLean, "Information systems success: the quest for the dependent variable," *Information Systems Research*, vol. 3, no. 1, pp. 60–95, 1992.

[9] J. L. Roldan and A. Leal, "A validation test of an adaptation of the DeLone and McLean's model in the Spanish EIS field," in *Proceedings of the BITWorld Conference*, June 2000.

[10] H.-Y. Jeong and Y.-H. Kim, "A system software quality model using DeLone & McLean model and ISO/IEC 9126," *International Journal of Digital Content Technology and its Applications*, vol. 6, no. 5, pp. 181–188, 2012.

[11] T. Algeddawy and H. Elmaraghy, "Manufacturing systems synthesis using knowledge discovery," *CIRP Annals*, vol. 60, no. 1, pp. 437–440, 2011.

[12] W. Zhou and S. Piramuthu, "Manufacturing with item-level RFID information: from macro to micro quality control," *International Journal of Production Economics*, vol. 135, no. 2, pp. 929–938, 2012.

[13] J. Y. Lee, D. Seo, B. Y. Song, and R. Gadh, "Visual and tangible interactions with physical and virtual objects using context-aware RFID," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3835–3845, 2010.

[14] B. Sun, Y. Xiao, C. C. Li, H.-H. Chen, and T. A. Yang, "Security co-existence of wireless sensor networks and RFID for pervasive computing," *Computer Communications*, vol. 31, no. 18, pp. 4294–4303, 2008.

[15] R. Van der Togt, P. J. M. Bakker, and M. W. M. Jaspers, "A framework for performance and data quality assessment of Radio Frequency IDentification (RFID) systems in health care settings," *Journal of Biomedical Informatics*, vol. 44, no. 2, pp. 372–383, 2011.

[16] T. C. Poon, K. L. Choy, H. K. H. Chow, H. C. W. Lau, F. T. S. Chan, and K. C. Ho, "A RFID case-based logistics resource management system for managing order-picking operations in warehouses," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8277–8301, 2009.

[17] S. Leimeister, J. M. Leimeister, U. Knebel, and H. Krcmar, "A cross-national comparison of perceived strategic importance of RFID for CIOs in Germany and Italy," *International Journal of Information Management*, vol. 29, no. 1, pp. 37–47, 2009.

[18] T. S. López, "RFID and sensor integration standards: state and future prospects," *Computer Standards & Interfaces*, vol. 33, pp. 207–213, 2011.

[19] H.-W. Jung, "Validating the external quality subcharacteristics of software products according to ISO/IEC 9126," *Computer Standards and Interfaces*, vol. 29, no. 6, pp. 653–661, 2007.

[20] C. K. M. Lee, W. Ho, G. T. S. Ho, and H. C. W. Lau, "Design and development of logistics workflow systems for demand management with RFID," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5428–5437, 2011.

[21] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: a ten-year update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9–30, 2003.

[22] W. H. Kruskal, "Ordinal measures of association," *Journal of the American Statistical Association*, vol. 53, pp. 814–861, 1958.

[23] E. Hartono, R. Santhanam, and C. W. Holsapple, "Factors that contribute to management support system success: an analysis of field studies," *Decision Support Systems*, vol. 43, no. 1, pp. 256–268, 2007.

[24] J. E. Hunter, F. L. Schmidt, and G. B. Jackson, *Meta Analysis: Cumulating Research Findings Across Studies*, Sage, Beverly Hills, Calif, USA, 1982.

[25] M. I. Hwang, J. C. Windsor, and A. Pryor, "Building a knowledge base for mis research: a meta-analysis of a system success model," *Information Resources Management Journal*, vol. 13, no. 2, pp. 26–32, 2000.