

Recent Advances in Green Communications and Networks

Lead Guest Editor: Yujin Lim

Guest Editors: Gianluigi Ferrari, Hideyuki Takahashi, and Rossana M. C. Andrade





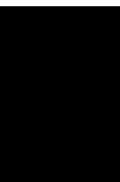
Recent Advances in Green Communications and Networks

Wireless Communications and Mobile Computing

**Recent Advances in Green
Communications and Networks**

Lead Guest Editor: Yujin Lim


Guest Editors: Gianluigi Ferrari, Hideyuki
Takahashi, and Rossana M. C. Andrade





Copyright © 2020 Hindawi Limited. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

































Zhipeng Cai , USA

Associate Editors

Ke Guan , China
Jaime Lloret , Spain
Maode Ma , Singapore

Academic Editors

Muhammad Inam Abbasi, Malaysia
Ghufran Ahmed , Pakistan
Hamza Mohammed Ridha Al-Khafaji ,
Iraq
Abdullah Alamoodi , Malaysia
Marica Amadeo, Italy
Sandhya Aneja, USA
Mohd Dilshad Ansari, India
Eva Antonino-Daviu , Spain
Mehmet Emin Aydin, United Kingdom
Parameshchhari B. D. , India
Kalapaveen Bagadi , India
Ashish Bagwari , India
Dr. Abdul Basit , Pakistan
Alessandro Bazzi , Italy
Zdenek Becvar , Czech Republic
Nabil Benamar , Morocco
Olivier Berder, France
Petros S. Bithas, Greece
Dario Bruneo , Italy
Jun Cai, Canada
Xuesong Cai, Denmark
Gerardo Canfora , Italy
Rolando Carrasco, United Kingdom
Vicente Casares-Giner , Spain
Brijesh Chaurasia, India
Lin Chen , France
Xianfu Chen , Finland
Hui Cheng , United Kingdom
Hsin-Hung Cho, Taiwan
Ernestina Cianca , Italy
Marta Cimitile , Italy
Riccardo Colella , Italy
Mario Collotta , Italy
Massimo Condoluci , Sweden
Antonino Crivello , Italy
Antonio De Domenico , France
Floriano De Rango , Italy

Antonio De la Oliva , Spain
Margot Deruyck, Belgium
Liang Dong , USA
Praveen Kumar Donta, Austria
Zhuojun Duan, USA
Mohammed El-Hajjar , United Kingdom
Oscar Esparza , Spain
Maria Fazio , Italy
Mauro Femminella , Italy
Manuel Fernandez-Veiga , Spain
Gianluigi Ferrari , Italy
Luca Foschini , Italy
Alexandros G. Fragkiadakis , Greece
Ivan Ganchev , Bulgaria
Óscar García, Spain
Manuel García Sánchez , Spain
L. J. García Villalba , Spain
Miguel Garcia-Pineda , Spain
Piedad Garrido , Spain
Michele Girolami, Italy
Mariusz Glabowski , Poland
Carles Gomez , Spain
Antonio Guerrieri , Italy
Barbara Guidi , Italy
Rami Hamdi, Qatar
Tao Han, USA
Sherief Hashima , Egypt
Mahmoud Hassaballah , Egypt
Yejun He , China
Yixin He, China
Andrej Hrovat , Slovenia
Chunqiang Hu , China
Xuexian Hu , China
Zhenghua Huang , China
Xiaohong Jiang , Japan
Vicente Julian , Spain
Rajesh Kaluri , India
Dimitrios Katsaros, Greece
Muhammad Asghar Khan, Pakistan
Rahim Khan , Pakistan
Ahmed Khattab, Egypt
Hasan Ali Khattak, Pakistan
Mario Kolberg , United Kingdom
Meet Kumari, India
Wen-Cheng Lai , Taiwan

Jose M. Lanza-Gutierrez, Spain
Pavlos I. Lazaridis , United Kingdom
Kim-Hung Le , Vietnam
Tuan Anh Le , United Kingdom
Xianfu Lei, China
Jianfeng Li , China
Xiangxue Li , China
Yaguang Lin , China
Zhi Lin , China
Liu Liu , China
Mingqian Liu , China
Zhi Liu, Japan
Miguel López-Benítez , United Kingdom
Chuanwen Luo , China
Lu Lv, China
Basem M. ElHalawany , Egypt
Imadeldin Mahgoub , USA
Rajesh Manoharan , India
Davide Mattera , Italy
Michael McGuire , Canada
Weizhi Meng , Denmark
Klaus Moessner , United Kingdom
Simone Morosi , Italy
Amrit Mukherjee, Czech Republic
Shahid Mumtaz , Portugal
Giovanni Nardini , Italy
Tuan M. Nguyen , Vietnam
Petros Nicolitidis , Greece
Rajendran Parthiban , Malaysia
Giovanni Pau , Italy
Matteo Petracca , Italy
Marco Picone , Italy
Daniele Pinchera , Italy
Giuseppe Piro , Italy
Javier Prieto , Spain
Umair Rafique, Finland
Maheswar Rajagopal , India
Sujan Rajbhandari , United Kingdom
Rajib Rana, Australia
Luca Reggiani , Italy
Daniel G. Reina , Spain
Bo Rong , Canada
Mangal Sain , Republic of Korea
Praneet Saurabh , India

Hans Schotten, Germany
Patrick Seeling , USA
Muhammad Shafiq , China
Zaffar Ahmed Shaikh , Pakistan
Vishal Sharma , United Kingdom
Kaize Shi , Australia
Chakchai So-In, Thailand
Enrique Stevens-Navarro , Mexico
Sangeetha Subbaraj , India
Tien-Wen Sung, Taiwan
Suhua Tang , Japan
Pan Tang , China
Pierre-Martin Tardif , Canada
Sreenath Reddy Thummaluru, India
Tran Trung Duy , Vietnam
Fan-Hsun Tseng, Taiwan
S Velliangiri , India
Quoc-Tuan Vien , United Kingdom
Enrico M. Vitucci , Italy
Shaohua Wan , China
Dawei Wang, China
Huaqun Wang , China
Pengfei Wang , China
Dapeng Wu , China
Huaming Wu , China
Ding Xu , China
YAN YAO , China
Jie Yang, USA
Long Yang , China
Qiang Ye , Canada
Changyan Yi , China
Ya-Ju Yu , Taiwan
Marat V. Yuldashev , Finland
Sherali Zeadally, USA
Hong-Hai Zhang, USA
Jiliang Zhang, China
Lei Zhang, Spain
Wence Zhang , China
Yushu Zhang, China
Kechen Zheng, China
Fuhui Zhou , USA
Meiling Zhu, United Kingdom
Zhengyu Zhu , China

Contents

LSTM-Based Traffic Load Balancing and Resource Allocation for an Edge System

Thembelihle Dlamini  and Sifiso Vilakati 


Research Article (15 pages), Article ID 8825396, Volume 2020 (2020)

Outage Probability Analysis of Decode-and-Forward Two-Way Relaying System with Energy Harvesting Relay

Hoang Van Toan, Tran Manh Hoang , and Le The Dung 


Research Article (13 pages), Article ID 8886487, Volume 2020 (2020)

Low-Cost Authentication Protocol for D2D Communication in m-Health with Trust Evaluation

Ana Paula G. Lopes and Paulo R. L. Gondim 


Research Article (16 pages), Article ID 8876807, Volume 2020 (2020)

Smart River Monitoring Using Wireless Sensor Networks

Kofi Sarpong Adu-Manu , Ferdinand Apietu Katsriku, Jamal-Deen Abdulai, and Felicia Engmann

Research Article (19 pages), Article ID 8897126, Volume 2020 (2020)

Censoring-Based Relay Transmission for Achieving Energy Efficiency in Cognitive Radio Networks

Li Feng , Jirong Sun, Hua Sun, and Shengju Tang

Research Article (14 pages), Article ID 5104605, Volume 2020 (2020)

Avoiding Spurious Retransmission over Flooding-Based Routing Protocol for Underwater Sensor Networks

Sungwon Lee , Yeongjoon Bae , Muhammad Toaha Raza Khan , Junho Seo , and Dongkyun Kim 


Research Article (9 pages), Article ID 8839541, Volume 2020 (2020)

A Simple and Energy-Efficient Flooding Scheme for Wireless Routing

Satoshi Yamazaki , Yu Abiko, and Hideki Mizuno


Research Article (9 pages), Article ID 8832602, Volume 2020 (2020)

Reducing the Energy Budget in WSN Using Time Series Models

Felicia Engmann , Ferdinand Apietu Katsriku, Jamal-Deen Abdulai, and Kofi Sarpong Adu-Manu


Research Article (15 pages), Article ID 8893064, Volume 2020 (2020)

The Hybrid Solar-RF Energy for Base Transceiver Stations

Cuong V. Nguyen, Minh T. Nguyen , Toan V. Quyen, Anh M. Le, and Linh H. Truong



Research Article (10 pages), Article ID 8875760, Volume 2020 (2020)

Association Game for Conflict Resolution between UEs and Small Cells

Jaesung Park 

Research Article (10 pages), Article ID 5801217, Volume 2020 (2020)

Analysis of Outage Probability and Throughput for Energy Harvesting Full-Duplex Decode-and-Forward Vehicle-to-Vehicle Relay System

Ba Cao Nguyen , Nguyen Nhu Thang, Tran Manh Hoang, and Le The Dung 
Research Article (10 pages), Article ID 3539450, Volume 2020 (2020)

Research Article

LSTM-Based Traffic Load Balancing and Resource Allocation for an Edge System

Thembelihle Dlamini ¹ and Sifiso Vilakati ²

¹Department of Electrical and Electronic Engineering, University of Eswatini, Kwaluseni, Eswatini

²Department of Statistics and Demography, University of Eswatini, Kwaluseni, Eswatini

Correspondence should be addressed to Thembelihle Dlamini; tldlamini@uniswa.sz

Received 10 July 2020; Revised 3 November 2020; Accepted 27 November 2020; Published 15 December 2020

Academic Editor: Gianluigi Ferrari

Copyright © 2020 Thembelihle Dlamini and Sifiso Vilakati. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The massive deployment of small cell Base Stations (SBSs) empowered with computing capabilities presents one of the most ingenious solutions adopted for 5G cellular networks towards meeting the foreseen data explosion and the ultralow latency demanded by mobile applications. This empowerment of SBSs with *Multi-access Edge Computing* (MEC) has emerged as a tentative solution to overcome the latency demands and bandwidth consumption required by mobile applications at the network edge. The MEC paradigm offers a limited amount of resources to support computation, thus mandating the use of intelligence mechanisms for resource allocation. The use of green energy for powering the network apparatuses (e.g., Base Stations (BSs), MEC servers) has attracted attention towards minimizing the carbon footprint and network operational costs. However, due to their high intermittency and unpredictability, the adoption of learning methods is a requisite. Towards intelligent edge system management, this paper proposes a Green-based Edge Network Management (GENM) algorithm, which is an online edge system management algorithm for enabling green-based load balancing in BSs and energy savings within the MEC server. The main goal is to minimize the overall energy consumption and guarantee the Quality of Service (QoS) within the network. To achieve this, the GENM algorithm performs dynamic management of BSs, autoscaling and reconfiguration of the computing resources, and on/off switching of the fast tunable laser drivers coupled with location-aware traffic scheduling in the MEC server. The obtained simulation results validate our analysis and demonstrate the superior performance of GENM compared to a benchmark algorithm.

1. Introduction

The foreseen *dense* deployment of BS empowered with computing capabilities in order to meet the ultralow latency demanded by mobile users raises concerns related to energy consumption. Apart from the fact that BS energy costs account for a large part of the operating expenses of MN operators, there are also increasing concerns regarding their environmental impact in terms of high carbon dioxide (CO₂) emissions. In an effort to minimize energy consumption and energy costs in 5G cellular networks within the MEC paradigm, this paper advocates for the integration of EH systems into network apparatuses and the use of container-based virtualization within computing platforms (i.e., MEC servers). The use of green energy mitigates the negative environmental

impact of MN and enables cost-saving for mobile operators in terms of lowering operational energy costs. The motivation towards green energy is due to the fact that current trends in battery and solar module costs show a reduction in prices. The benefits of container-based virtualization is the reduction in energy drained in the computing platform due to their lower overheads when compared with VM [1, 2]. For a qualitative comparison of different virtualization techniques, interested readers are referred to [1].

In this paper, a group of EH BS placed in proximity to a EH-MEC server is considered as an *edge system*. The EH-MEC server manages the BS system, deciding upon the allocation of transmission resources and also handling the computing and communication processes. In general, renewable energy systems are dimensioned to guarantee

the autonomous operation of the edge system. Thus, it is desirable that the utilization of green energy be made one of the performance metrics when designing load-balancing strategies [3, 4], instead of the network impact [5, 6]. As a result, the green-based load balancing is a promising technique for optimizing MEC performance since it exploits the spatial diversity of the available green energy to reshape the network load among the BSs [4]. In this case, MN can prioritize the utilization of BS with sufficient green energy to serve more traffic while reducing the traffic loads for those BS consuming on-grid power. For instance, in the MEC server, a trade-off between the green energy utilization and the amount of workload that can be computed locally should be carefully evaluated.

In this regard, it is worth noting that the energy consumption within the virtualized computing platform is due to (i) the active computing resources, i.e., VM or containers [7–9], and (ii) the network communications, communication related to transmission drivers [10, 11], and the intracommunications [12]. In order to alleviate this, this paper assumes that the container-based virtualization be deployed in the MEC server as containers are lightweight, i.e., demand less memory space, have shorter start-up time, and offer software portability. At each time instance, the containers are provisioned based on the forecasted server workloads, a technique referred to as *autoscaling*. In addition, the transmission drivers used for data transfers within the server are tuned by taking into account the mobile user's location. The server is also capable of caching the frequently requested contents locally.

In densely deployed BS, the energy drained is due to the always-on design approach [5, 6, 13], yet traffic load varies during the day, e.g., the demand is low during the night. Therefore, in order to intelligently manage the BS towards minimizing the energy consumption, the green-based load balancing technique is employed, i.e., BS sleep modes are enabled in some BS using green energy as a performance metric.

1.1. Paper Contributions. This paper considers an energy cost model that takes into account the computing, caching and communication processes within the MEC server, and transmission-related energy consumption in BS. Here, the GENM algorithm is proposed for enabling *green-based* traffic load balancing, i.e., the BS are dynamically switched on/off based on their harvested energy, autoscaling, and reconfiguring the computing resources, and the tuning of transmission drivers. This entails using a minimum number of optical drivers for real-time data transfers, over a short-term horizon. In order to solve the energy consumption minimization problem, the GENM algorithm performs online supervisory control, utilizing the learned traffic load and the harvested energy patterns. Then, the output is utilized within a LLC policy [14] to obtain the optimal system control actions that yield the desired energy savings that guarantee the required QoS.

This work is an extension of [11], where energy savings and QoS guarantee were considered only within a virtualized computing platform placed in proximity to a cluster of BS. In [11], the role of the MEC server is to handle the offloaded computational workload only, which means that the energy cost model lacks the consideration of the BS management procedures, caching process, and the use of containers.

The summary of contributions are listed as follows:

- (i) The use of container-based virtualization is introduced as they are lightweight, i.e., demand less memory space, have shorter start-up time, and offer software portability
- (ii) The proposed GENM algorithm, which is an online edge management system, makes use of predictive optimization, specifically using the LLC, where green-based load balancing, container provisioning, and the tuning of the transmission drivers is performed based on the learned information
- (iii) The numerical results, obtained with real-world harvested energy and traffic load traces, show that the proposed optimization strategy is able to efficiently manage the edge network resources in order to minimize the energy drained under the guidance of the intelligent online-based resource manager and the energy-saving procedures

In order to achieve these, the remainder of the paper is organized as follows: Section 2 describes the related work. Section 3 explains the system model. In Section 4, the design and the implementation of the online algorithm are presented. Simulated results are discussed in Section 5. Lastly, the work is concluded in Section 6.

2. Related Work

2.1. Methods for Load Balancing in MN. Load balancing has been studied towards data center management whereby the data center servers employ temporal dependency strategies, i.e., the servers are turned on/off depending on the arrival rates of workloads. This significantly differs from our considered problem as we consider load balancing in SBSs powered by green energy. Towards load balancing, the dynamic BS switching on/off strategies have been used. However, this may have an impact on the network due to the load that is offloaded to the neighboring BS(s). To avoid this, the BS to be switched off is carefully identified within the BS cluster. In [5, 6], the *network impact* is used to identify the BS to be switched off, one at a time, with no significant network performance degradation. Taking into account daily traffic load variation, strategies for opportunistic utilization of the unexploited third-party small cell (SC) capacity is exploited towards energy savings in [15], in order to enable the switching off of some BSs. Here, an offloading mechanism is introduced, where the operators lease the capacity of a SC network owned by a third party in order to switch off their BS (Macro BS) and maximize their energy efficiency, when the traffic demand is low.

The use of green energy as a performance metric has been explored within the Radio Access Network (RAN) [3, 16]. Along the lines of MN softwarization, a distributed user association scheme that makes use of the SoftRAN concept for traffic load balancing via the RAN Controller (RANC) is proposed in [3]. Here, the user association algorithm runs on the RANC, and the users report their downlink data rates via the associated BS to the RANC, where the traffic loads

from individual users and BS are measured. The algorithm enhances the network performance by reducing the average traffic delivery latency in BS as well as to reduce the on-grid power consumption by optimizing the green energy usage. Then, the authors in [16] proposed to optimize the utilization of green energy for cellular networks by optimizing the BS transmission power. The proposed scheme achieves significant on-grid power savings by scheduling the green energy consumption along the time domain for individual BS and balancing the green energy consumption among the BSs.

Along the lines of MEC, the authors in [4] proposed a framework for jointly performing load balancing, admission control, and energy purchase within a network of EH-powered BS with the goal of minimizing the computation delay and data traffic drops (i.e., increasing the locally computed workloads). This work uses green energy as a performance metric. To solve this problem, an online and distributed algorithm is proposed leveraging the Lyapunov optimization with the perturbation technique. Here, the algorithm makes the traffic load decisions without forecasting the future traffic load and harvested energy. In contrast, the work presented in this paper considers the short-term future traffic load and the harvested energy for decision making. Then, in our previous work [6], a supervisory online control algorithm that makes use of clustering and the network impact metric towards load balancing in MN is proposed. Here, the BSs are empowered with computation capabilities (with VM as computing resources), the LSTM neural network is used for forecasting, and the LLC policy handles foresighted optimization. Even though these works perform load balancing, the problem and scenario considered in this paper are different. Here, a MEC server manages the SBSs powered by green energy. Similar to [11], forecasting and foresighted optimization are used for edge system management.

Load balancing that follows the energy routing, i.e., more computational workload is offloaded to where more energy is available, is presented in [17]. To handle spatial uneven computation workloads experienced by the MEC-enabled BSs, the authors proposed a peer offloading scheme. Here, the BSs share their computing resources and energy costs.

2.2. Methods for Energy Saving within Computing Platforms.

Green computing over data centers is an emerging paradigm that aims at performing the dynamic energy-saving management of data center infrastructures. Here, procedures for the dynamic on/off switching of servers have been proposed as a way of minimizing energy consumption in computing platforms. A novel postdecision state-based learning algorithm for server provisioning at the network edge is presented in [18]. This work incorporates green energy. At the beginning of the time slot, the servers are consolidated, i.e., the number of turned on physical servers are minimized, using the learned optimal policy for dynamic workload offloading and the autoscaling (or right-sizing). Then, in our previous works [6, 11, 19], VM *soft-scaling* (i.e., the reduction of computing resources per time instance) is employed towards energy saving in virtualized platforms either energized by only renewable energy or hybrid supplies (solar and power grid). This is achieved by forecasting the traffic load and har-

vested energy and then employing foresighted optimization to obtain the system control inputs. The work of [20] uses an iterative algorithm to obtain the number of computing resources (VMs) to be provisioned within a node that transmits to clients wireless. Then, the work of [21] considers a vehicular scenario where vehicles connect wireless to fog nodes and then develop an adaptive scheduler, which computes on-the-fly the solutions of both the resource reconfiguration and consolidation problems. For this purpose, the primal-dual algorithm is used.

In computing platforms, computation offloading strategies can be jointly exploited together with delay constraints towards energy savings. The authors in [22] proposed an offloading policy to find the optimal place where to offload and the amount of offloaded task data. In this work, the time taken for processing the offloaded task is reduced, at the same time consuming less energy. Then, in [23], an efficient scheduling for latency-sensitive applications is proposed towards energy and response time minimization. The achieved results show a reduction in delay and network usage and the energy consumption. In addition, the works of [24] jointly optimize the computing and communication resources, taking into account the local task execution delay and transmission delay. To meet the task delay requirements, in [25], the heterogeneous clouds, i.e., edge and remote cloud, are coordinated. Here, different policies are employed in the clouds. In this, the edge cloud handles tasks with loose delay bounds and drops tasks with stringent delay bounds when the traffic load is heavy.

Towards minimizing energy consumption induced by communication activities within a computing node, the idea of tuning transmission drivers, as one of the energy-saving strategies within the MN infrastructure, is first conceived in [10, 26] where a computing node (router) is considered. Here, it is observed that having the least number of data transmission drivers (*fast* tunable lasers) can yield a significant amount of energy savings. Motivated by the aforementioned works, within the MEC paradigm, the authors in [11] put forward a traffic engineering- and MEC LS-based algorithm that uses a location-aware procedure for provisioning the transmission drivers for data transfer towards target BS. Here, the MEC LS API is employed for retrieving the UE's location and then passing the information to the authorized applications within the MEC platform, for decision-making.

Overall, these works numerically analyze and test the energy performance of some state-of-the-art schedulers for computing platforms, but do not attempt to optimize it through the dynamic joint scaling of the available communication-plus-computing resources. The joint analysis of the computing-plus-communication energy consumption within the MEC paradigm is still an open research topic.

2.3. Methods for Guaranteeing Quality of Service and Enabling Energy Savings (within the MEC Paradigm).

The mobile operator's goal is to provide QoS Internet services for large populations of clients, while minimizing the overall computing-plus-communication energy consumption. Hence, a trade-off is required between QoS and energy savings. Future MNs are expected to learn the diverse characteristics of user behavior, as well as renewable energy variations, in

order to autonomously determine good system configurations. Towards this goal, online forecasting using ML techniques and the LLC method can yield the desired system behavior when taking into account the environmental inputs, i.e., BS traffic load, server workloads, and energy to be harvested. Next, the mathematical tools that are used in this research work are reviewed, namely, the LLC method [14, 27, 28] and LSTM neural networks [29, 30].

The LLC has been used in [27] to address a resource provision problem within virtualized environments. The optimization problem is posed as a profit maximization problem under uncertainty and the LLC formulation models the cost of control. To address the uncertainty over the workload arrival, the Kalman filter is used. Then, in [31], an online supervisory control scheme based on LLC policies is proposed. Here, after the occurrence of an event, the next control action is determined by estimating the system behavior a few steps into the future using the currently available information as inputs. The control action exploration is performed using a search tree assuming that the controller knows all future possible states of the process over the prediction horizon. An online control framework for resource management in switching hybrid systems is proposed in [14], where the system's control inputs are finite. The relevant parameters of the operating environment, e.g., workload arrival, are estimated and then used by the system to forecast future behavior over a look-ahead horizon. From this, the controller optimizes the predicted system behavior following the specified QoS through the selection of the system controls. In [6, 19], a LLC-based supervisory algorithm is proposed to obtain the system control actions yielding the desired trade-off between energy consumption and QoS. Here, the traffic load and harvested energy are forecasted and then used as input in the algorithm. The BS are densely deployed in [6], and each BS is empowered with computation capabilities. Then, in [19], a remote site powered by only green energy is considered.

LSTM can be used for multistep time series forecasting as it is able to handle the long-term dependencies due to its inherent capability of storing past information and then recalling it. The forecasting method is utilized in [6, 19] within an LLC-based algorithm to obtain the system control actions yielding the desired trade-off between energy consumption and QoS. The application of the LSTM network is extended to include ITS in [32]. A new ITS edge analytics architecture that makes use of deep learning techniques that run either on the mobile devices or on the intravehicle processors for data analytics is presented. A combination of LSTM networks and deep CNN is adopted, i.e., CNN-LSTM network, for path selection in autonomous vehicles, whereby the CNN is used for feature extraction, and then, the extracted information is fed into LSTM networks for driving path selection. Forecasting server workloads using the LSTM network can be beneficial for dynamic resource scaling and power consumption in cloud computing data centers. In [33], a forecasting model using the LSTM network for predicting future data center workloads is proposed, and then, the results are fed into the resource manager for decision-making, which involves either scaling up or down the computing resources (servers in this case).

3. System Model

In line with ETSI-proposed MEC deployment scenarios discussed in [34], the considered network scenario is illustrated in Figure 1 where the proposed model is cache-enabled, TCP/IP offload capable (i.e., enables *partial* offloading in the server's NIC such as checksum computation [35]). The virtualized MEC server is assumed to be hosting C containers deployed at an aggregation point, which is in proximity to a cluster of N BS from the same MN operator. The BS are interconnected to the MEC server for computation workload offloading. Each network apparatus (BS, MEC server) is mainly powered by renewable energy harvested from wind and solar radiation, and it is equipped with an EB for energy storage. In this case, energy can only be purchased from the grid supply to supplement the renewable energy supplies. The BS coverage areas overlap so that green-based load balancing is possible. The EM is an entity responsible for selecting the appropriate energy source to fulfill the EB and also for monitoring the energy level of the EB. Then, the electromechanical switch (SW) aggregates the energy sources to fulfill the EB level. In the MEC server, there is the presence of a virtualized access control router which acts as an access gateway for admission control, responsible for local and remote routing, and it is locally hosted as an application. Also, the MEC platform is able to track user location via the MEC Location Service Application Programmable Interface (LS API). Lastly, a discrete-time model is considered whereby time is discretized as $t = 1, 2, \dots$ and each time slot t has a fixed duration $\tau = 30$. The list of symbols that are used in the paper is reported in Table 1.

3.1. Communication Traffic and Energy Consumption. From a networking perspective, the understanding and characterization of energy consumption within the MN can pave the way towards more efficient and user-oriented networking solutions. This can be achieved through the use of historical mobile traffic traces such as Call Detail Records (CDRs) obtained from mobile operators, specifically in the EPC network. Due to the difficulties in obtaining relevant open-source datasets containing computing requests, real MN traffic load traces obtained from the TIM network (availed through the Big Data Challenge [36]) are used to emulate the computational load. In order to understand the daily traffic load patterns, the clustering algorithm X-means [37] has been applied to classify the load profiles into several categories. Here, each BS n is assumed to have a related load profile $L_n(t)$ which is picked at random as one of the four clusters in Figure 2. In addition, it is assumed that $L_n(t)$ consists of 80% delay sensitive workloads $\gamma_n(t)$ and the remainder is delay-tolerant workloads. The total aggregate delay-sensitive workload per time instance is $\xi(t) = \sum_{n=1}^N \gamma_n(t)$.

The virtualized router in the MEC server of Figure 1 determines the amount of workload that can be accepted by the input buffer at slot t , and the aggregated (or admitted) workload is denoted by $L_{in}(t) \in [0, L_{in}^{\max}]$ (measured in [Mbits]). L_{in}^{\max} is the maximum input buffer size. In addition, it is assumed that the input/output (I/O) queues of the MEC server are loss-free, and they implement the First-In First-

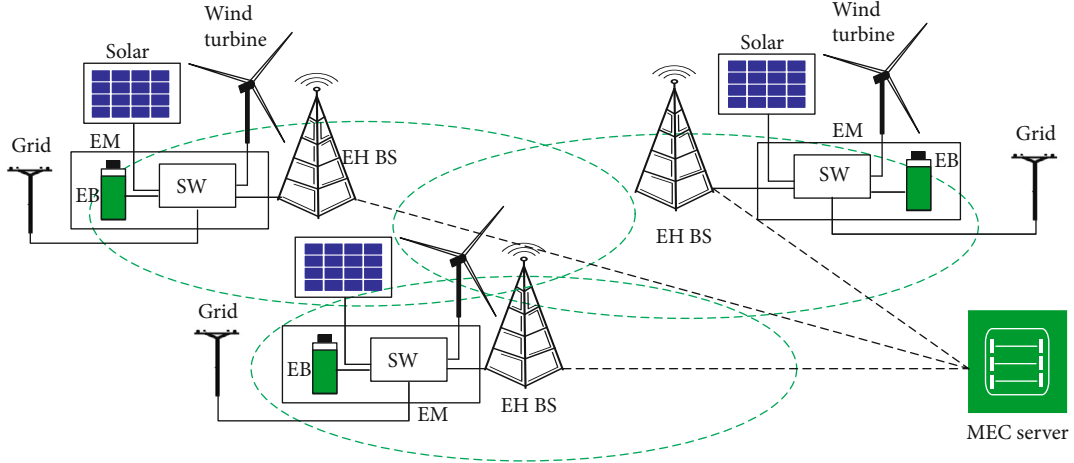


FIGURE 1: Edge system energized by hybrid energy sources: on-grid power and green energy (solar and wind).

Out (FIFO) service discipline, thus $L_{in}(t) = L_{out}(t)$, where $L_{out}(t)$ is the amount of the aggregate computed workload at the output buffer.

The total energy consumption ([J]) for the edge system at time slot t is formulated as follows, inspired by [11, 12]:

$$\theta_{EDGE}(t) = \theta_{COMM}(t) + \theta_{MEC}(t), \quad (1)$$

where $\theta_{COMM}(t)$ is the energy consumption term induced by all BS communications and $\theta_{MEC}(t)$ is the energy consumption term induced by the MEC server's computing, caching, and communication processes.

3.1.1. BS Energy Cost. The overall energy consumption within the coverage area is defined as the sum of all the BS components:

$$\theta_{COMM}(t) = \sum_{n=1}^N \theta_{BS,n}(t) = \sum_{n=1}^N (\delta_n(t)\theta_0 + \theta_{load,n}(t)), \quad (2)$$

where $\delta_n(t) \in \{0, 1\}$ is the BS n switching status indicator (1 for *active mode* and 0 for *power saving mode*); θ_0 is a constant value (load independent) representing the operation energy which includes baseband processing, radio frequency power expenditures, etc.; $\theta_{load,n}(t)$ is the load-dependent BS transmission power to the served users that guarantees low latency at the edge. It is obtained by using the transmission model in [38].

3.1.2. MEC Energy Cost. The energy drained due to the computing, caching, and communication processes is defined as:

$$\begin{aligned} \theta_{MEC}(t) = & \theta_{CNT}(t) + \theta_{SWT}(t) + \theta_{OFF}(t) + \theta_{LNK}(t) + \theta_{DR}(t) \\ & + \theta_{CC}(t), \end{aligned} \quad (3)$$

where $\theta_{CNT}(t)$ is the energy drained due to the active containers, w.r.t CPU utilization, and $\theta_{SWT}(t)$ is the energy drained due to containers adapting to new processing rates $f_c(t) \in [f_0, f_{max}]$ [(Mbit/s)]. The term f_0 is the zero speed of the container, e.g., deep sleep or shutdown, and f_{max} is the maximum available processing rate for container c . It is

worth noting that actual containers are generally instantiated atop physical computing cores which offer only a finite set of processing speeds. The term $\theta_{OFF}(t)$ is the energy induced by the TCP/IP offload on the NIC, and $\theta_{LNK}(t)$ is the energy drained due to the virtual-links (to-and-from containers) communication cost. Then, $\theta_{DR}(t)$ is the amount of energy consumed by the active transmission drivers, and $\theta_{CC}(t)$ is the total energy cost incurred by the content caching process.

In this regard, it is assumed that real-time processing of computation workloads are performed in parallel over the containers interconnected by a rate-adaptive Virtual LAN (VLAN). In addition, it is also assumed that the CPU frequency is fixed at each user and may vary over users. The amount of energy consumed by the CPU is related to the provisioned computing resources, i.e., the CPU share allocated to each container, per time instance t , named $C(t) \leq C$, index by c . Thus, $\theta_{CNT}(t)$ is defined as [19]:

$$\theta_{CNT}(t) = \sum_{c=1}^{C(t)} \theta_{idle,c}(t) + \psi_c(t)(\theta_{max,c}(t) - \theta_{idle,c}(t)), \quad (4)$$

where $\theta_{idle,c}(t)$ represents the *static* energy drained by container c in the idle state, $\psi_c(t) = (f_c(t)/f_{max})^2$ is the utilization function of container c [28], and $\theta_{max,c}(t)$ is the *maximum* energy that container c can consume. The quantity $\psi_c(t)(\theta_{max,c}(t) - \theta_{idle,c}(t))$ represents the *dynamic* energy component of container c .

The intelligent resource manager implements a suitable frequency-scaling policy in real-time, in order to allow the containers to scale up/down their processing rates $f_c(t)$ at the minimum cost. In this regard, it should be noted that switching from the processing frequency $f_c(t-1)$ (the processing rate at the $(t-1)$ time instance) to the next processing frequency $f_c(t)$ entails an energy cost, $\theta_{SWT}(t)$. This depends on the absolute processing rate gap $|f_c(t) - f_c(t-1)|$; thus, $\theta_{SWT}(t)$ is defined as:

$$\theta_{SWT}(t) = \sum_{c=1}^{C(t)} z_e (f_c(t) - f_c(t-1))^2, \quad (5)$$

TABLE 1: Notation: list of symbols used in the analysis.

Symbol	Description
Input parameters	
C	Maximum number of containers hosted by the MEC server, indexed by c
N	Number of BS, indexed by n
$\xi(t)$	Aggregate computational workload
τ	Time slot duration
$L_{in}(t)$	Amount of aggregate workload at the input buffer
$L_{out}(t)$	Amount of aggregate workload at the output buffer
$L_{out}^{\max}(t), L_{in}^{\max}(t)$	Workload buffers maximum capacity
f_{\max}	Maximum processing rate for container c
$\theta_{idle,c}(t)$	Static energy consumed by container c in the idle state
$\theta_{\max,c}(t)$	Maximum energy consumed by container c at maximum processing rate
z_e	Per-container reconfiguration cost caused by a unit-size frequency switching
$\lambda_c(t)$	Workload fraction to be computed by the c -th container
λ_{\max}	Maximum computation load per container
Δ	Maximum per-slot and per-container allowed processing time
$\theta_{idle}^{NIC}(t)$	Energy drained by the NIC when powered, with no data transfer
M	Maximum number of multiple fast tunable lasers
β_{\max}	Maximum energy buffer capacity
β_{up}, β_{low}	Upper and lower energy buffer thresholds
Variables	
$\theta_{COMM}(t)$	Total BS energy consumption at time slot t
$\theta_{MEC}(t)$	Server's energy consumption at time slot t
$\theta_{CNT}(t)$	Energy drained due to the active containers, w.r.t CPU utilization, at time slot t
$\theta_{SC}(t)$	Energy drained due to container switching the processing rates at time slot t
$\theta_{OFF}(t)$	Energy induced by the TOE at t
$\theta_{LNK}(t)$	Energy drained due to the virtual-link communication cost at time slot t
$\theta_{DR}(t)$	Energy drained due to the number of active transmission drivers at time slot t
$\theta_{CC}(t)$	Total energy cost incurred by the content caching process
$C(t)$	Number of containers to be active in time slot t
$f_c(t)$	Instantaneous processing rate
$\psi_c(t)$	Load-dependent factor
$r_c(t)$	c -th virtual link communication rate at slot t
$\zeta_n(t)$	BS switching status indicator at t
$\theta_{\max}^{NIC}(t)$	Maximum energy drained by the TOE at t
$\chi_t(t)$	The expected processing time

TABLE 1: Continued.

Symbol	Description
$M(t)$	Number of active transmission drivers at t
$b(t)$	Energy buffer level in slot t
$H(t)$	Harvested energy profile in slot t
$E(t)$	Purchased grid energy in slot t

where z_e is the per-container reconfiguration cost caused by a unit-size frequency switching. Typically, z_e is limited to a few hundreds of millijoule per square megahertz.

Before proceeding, it is worth noting the following: at the beginning of the time slot t , the online algorithm adaptively allocates the available resources and then determine the containers that are demanded, $C(t)$; the size of the workload allocated to the container c , denoted by $\lambda_c(t)$; and $f_c(t)$ for container c that will yield the desired or expected processing time, $\chi_c(t) = \lambda_c(t)/f_c(t)$. $\chi_c(t) \leq \Delta$, where Δ is the maximum per-slot and per-container processing time ([s]). Note that $L_{in}(t) = \sum_{c=1}^{C(t)} \lambda_c(t)$ is the amount of computational workload admitted in the MEC server, by the router. The amount of the workload to be admitted per slot shall be decided at the beginning of each time slot depending on the forecasted green energy, grid power to be purchased, and the expected computational workloads $\hat{L}_{in}(t)$. Moreover, virtualization technologies specify the minimum and maximum amount of resources that can be allocated per container [40]; thus, the maximum amount is denoted by λ_{\max} . Lastly, the container(s) provisioning and workload allocation is discussed in Section 4.2, Remark 1, and $f_c(t) = \Delta \lambda_c(t)/\Delta$.

By implementing a TCP Offload Engine (TOE) in high-speed computing environments, some TCP/IP processing is offloaded to the network adapter for the purpose of reducing the CPU utilization. To obtain the energy cost incurred, the performance measure for the Broadcom (Fibre) 10 Gbps NIC [35] is considered here as an example of a TCP/IP offload-capable device. Note that $\theta_{OFF}(t)$ is data volume-dependent and it is obtained as:

$$\theta_{OFF}(t) = \zeta(t) \theta_{idle}^{NIC}(t) + \theta_{\max}^{NIC}(t), \quad (6)$$

where $\theta_{idle}^{NIC}(t) > 0$ is the energy drained by the TOE when powered, with all links connected without any data transfer. This provides an opportunity for switching off the network adapter if there is no data transfer, making the energy drained to be zero. For this, $\zeta(t) = (0, 1)$ is the switching status indicator (1 for active state and 0 for idle state) and $\theta_{\max}^{NIC}(t) = (g(t) \cdot L_{in}(t))/\eta$ is the maximum energy drained, where $g(t)$ is a fractional value representing the amount of load computed in the network adapter and η is the NIC best throughput performance, hereby obtained as a fixed value measured in [Gbit/].

In order to keep the transmission delays from (to) the scheduler to (from) the connected containers at a minimum value, it is assumed that each container c communicates with the resource scheduler through a dedicated reliable link that

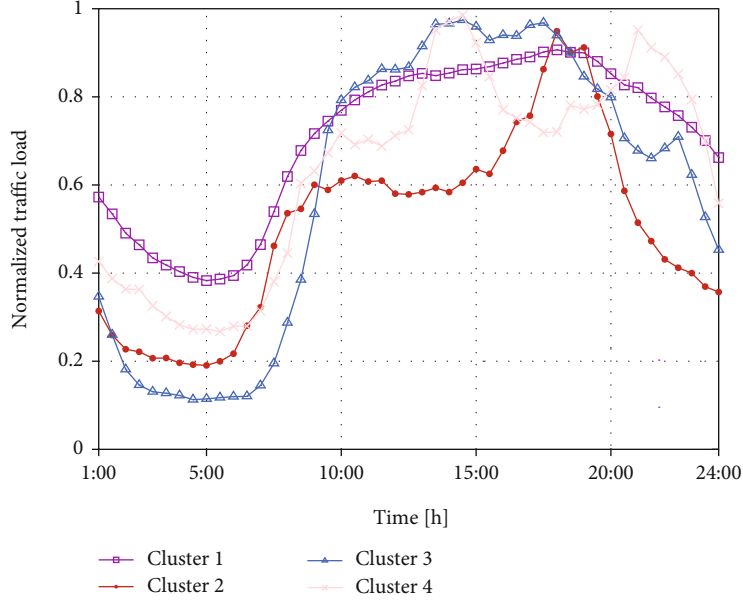


FIGURE 2: Normalized BS traffic load behavior represented as clusters. The data from [39] has been split into four representative clusters.

operates at the transmission rate of $r_c(t)$ [(bit/s)]. Thus, the energy needed for sustaining the two-way c^{th} link is defined, inspired by [41]:

$$\theta_{\text{LNK}}(t) = 2 \sum_{c=1}^{C(t)} P_c(r_c(t)) (\lambda_c(t)/r_c(t)), \quad (7)$$

where $P_c(r_c(t)) = S_c(2^{r_c(t)/W_c} - 1)$ is the power drained by the c^{th} communication link and $S_c = (W_c \times N_0^{(c)})/g_c$, $N_0^{(c)}$ (W/Hz) is the noise spectral power density, W_c is the bandwidth, and g_c is the (nonnegative) gain of the c^{th} link. In practical application scenarios, the maximum per-slot communication rate within the intra-VLAN is generally limited up to an assigned value r_{max} . Thus, the following hard constraint must hold: $\sum_{c=1}^{C(t)} r_c(t) \leq r_{\text{max}}$.

In this regard, a two-way per-task execution delay is considered. Here, there is a total of $c = \{1, \dots, C(t)\}$ link connection delays, each denoted by $\rho_c(t) = \lambda_c(t)/r_c(t)$, and $\chi_c(t) \leq \Delta$ where Δ is the server's response time, i.e., the maximum time allowed for processing the total computation load, and it is fixed in advance regardless of the task size allocated to container c . Since parallel real-time processing is assumed in this work, the overall communication equates to $2\rho_c(t) + \Delta$. Therefore, the hard per-task delay constraint on the computation time is: $\max\{2\rho_c(t) + \Delta\} = \tau_{\text{max}}$, where τ_{max} is the maximum tolerable delay, which is fixed in advance.

Edge distributed devices utilize low-level signaling for information sharing. Thus, edge computing systems receive information from mobile devices within the local access network to discover their location. In return, for every *client* who offloaded their task into the MEC server associated with the radio nodes, i.e., BS, its location and the computation result are known through the LS (which is a service that supports UE location retrieval mechanism, and then passing the infor-

mation to the authorized applications within the server), thus enabling the location-aware traffic routing and obtaining the number of transmission drivers to be active for data transfers. The term $\theta_{\text{DR}}(t)$ depends on the number of active laser (optical) drivers, named $M(t) \leq M$, where M is the total number of drivers that are required for transferring $\ell_m(t) \in L_{\text{out}}(t)$ in time slot t ($\ell_m(t)$ is the downlink traffic volume ([bits]) of the driver at slot t). $L_{\text{out}}(t)$ is accumulated over a fixed period of time to form a *batch* at the output buffer. This means that a large number of drivers yield large transmission speed while at the same time resulting in high energy consumption [26]. Therefore, the energy consumption can be minimized by launching an optimal number of drivers for the data transfer.

The energy drained during the data transmission process consists of the following: a constant energy for utilizing each fast tunable driver denoted by $d_m(t)$ ([J/s]), the target transmission rate r_0 [bits/s], and $L_{\text{out}}(t)$. Thus, the energy is inspired by [11]:

$$\theta_{\text{DR}}(t) = \sum_{m=1}^{M(t)} \frac{d_m(t) l_m(t)}{r_0}, \quad (8)$$

where the parameter $M(t)$ is obtained using the total number of target BS as $M(t) = \lceil 1/u \cdot ((\omega(t) + 1)/(\omega(t)))^2 \rceil$, where $\omega(t) = \sqrt{\rho/(\sigma N_{\text{BS}}(t))} \in (0, 1]$ is a controllable factor that determines the delay constraint of optical networks, σ ([ms]) is the reconfiguration cost for tuning the transceivers, $N_{\text{BS}}(t)$ is an integer value representing the total number of target BS at time slot t , and ρ is the number of time slots at which the computed workload is accumulated at the output buffer. Thus, the terms σ and ρ are fixed values, and $L_{\text{out}}(t)$ is equally distributed over the $M(t)$ drivers.

The MEC server is able to cache contents from the Internet and store the contents closer to mobile users. The caching process also contributes to the energy consumption in the server. The caching process is restricted to only viral content.

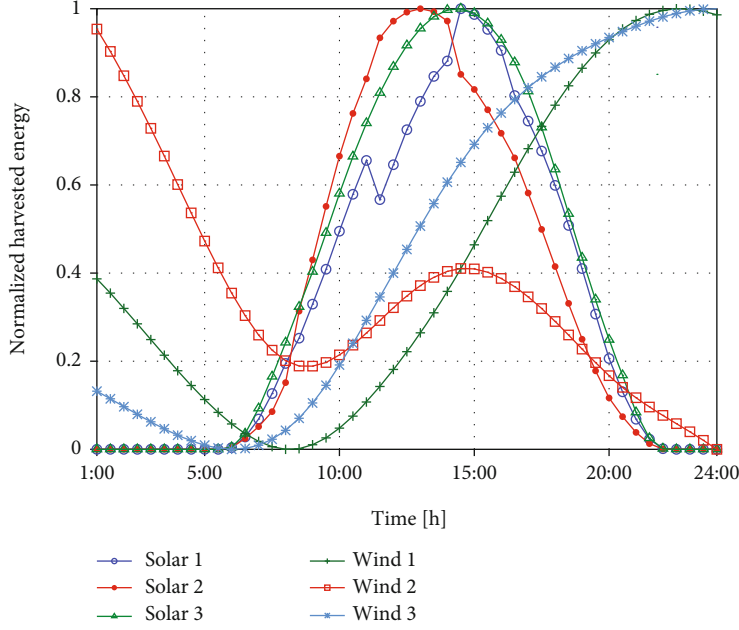


FIGURE 3: Example traces for harvested solar traces and wind traces from [43].

For example, when a video becomes viral, users watching it share and talk about that video, which will be then requested by other users after a response time. Taking into account the Internet users' response time $\bar{\lambda}(t)$, this epidemic behavior can be modelled by the self-excited Hawkes condition Poisson process described in [42]: $\bar{\lambda}(t) = V(t) + \sum_{t_i \leq t} \Omega_i k_i(t - t_i)$, where $k_i(t)$ is the response time function and Ω_i is the number of potential viewers who will be influenced after t_i , which is the time when the user i shared the video. The term $V(t)$ is added as a component of the model to capture the views that are not triggered by the epidemic effect. The energy consumption is mainly contributed by content caching and data transmission processes, as such $\theta_{CC}(t)$ is defined as:

$$\theta_{CC}(t) = \bar{\lambda}(t) (\theta_{TR}(t) + \theta_{CACHE}(t)), \quad (9)$$

where $\theta_{TR}(t)$ is the power consumption due to transmission and $\theta_{CACHE}(t)$ is the power consumption contributed by the caching process.

3.2. Energy Patterns and Storage. The rechargeable energy storage device is characterized by its finite energy storage capacity b_{max} . At each time instance, the energy level reports are pushed from the BS sites to the MEC server. Thus, the EB level $b(t)$ is known, enabling the provisioning of the required computation and communication resources, i.e., the required containers, transmission drivers, and BSs to be active. In this paper, the amount of harvested energy $H(t)$, per site BS site, in time slot t is obtained from open-source solar and wind traces from a farm located in Belgium [43] (see Figure 3 above). The data in the dataset matches our time slot duration (30). The dataset is the result of daily environmental records for a place assumed to be free from surrounding obstructions (e.g., buildings, shades).

The harvested energy $H(t)$ is obtained by picking a day at random in the dataset and associating it with one site. Here, the wind energy is selected as a source during the solar energy off-peak periods. The available EB level $b(t+1)$ located at the BS site (BS n) or computing platform evolves according to the following dynamics:

$$b(t+1) = \min \{b(t) + H(t) - \theta_{site}(t) - a(t) + E(t), b_{max}\}, \quad (10)$$

where $b(t)$ is the energy level in the battery at the beginning of the time slot t and $\theta_{site}(t)$ represents either $\theta_{BS,n}(t)$, the BS energy consumption of the communication site, or $\theta_{MEC}(t)$, the energy drained at the computing platform, over time slot t , see Eqs. (2) and (3). $a(t)$ is leakage energy, and $E(t) \geq 0$ is the amount of energy purchased from the power grid. It is worth noting that $b(t)$ is updated at the beginning of time slot t , whereas $H(t)$, $\theta_{BS,n}(t)$, and $\theta_{MEC}(t)$ are only known at the end of it. Thus, the energy constraint at the computing site must be satisfied for every time slot: $\theta_{MEC}(t) \leq b(t)$.

For decision-making in the GENM application, the received EB level reports are compared with the following thresholds: b_{low} and b_{up} , respectively, termed the lower and the upper energy threshold with $0 < b_{low} < b_{up} < b_{max}$. b_{up} corresponds to the desired energy buffer level at the BS site or computing site, and b_{low} is the lowest EB level that any site should ever reach. If $b(t) < b_{low}$, then BS n or the computing site is said to be *energy deficient*. The suitable energy source at each time slot t is selected based on the forecast expectations, i.e., the expected harvested energy $\hat{H}(t)$. If $\hat{H}(t)$ is enough to reach b_{up} , no energy purchase is needed. Otherwise, the remaining amount up to b_{up} , i.e., $E(t) = b_{up} - b(t)$, is purchased from the electrical grid. Our optimization framework

in Section 4.1 makes sure that $b(t)$ never falls below b_{low} and guarantees that b_{up} is reached at every time slot.

4. Problem Formulation

In this section, the optimization problem is formulated to obtain reduced energy consumption through short-term traffic load and harvested solar energy forecasting along with energy management procedures. The optimization problem is defined in Section 4.1, and the edge system management procedures are presented in Section 4.2.

4.1. Optimization Problem. Our objective is to improve the overall energy savings of the edge system through BS power saving modes (i.e., green-based traffic load balancing), autoscaling of containers, contents caching, and tuning of the transmission drivers, and also to guarantee the QoS within the network. Note that at the end of each time slot, the EB states are updated depending on the harvested energy and the consumed energy, thereby linking per-time slot problems across time.

To achieve our objective, two cost functions are defined, one captures the edge system energy consumption and the other handles the QoS. This is defined as $(F1)\theta_{\text{EDGE}}(t)$, which weighs the energy consumption due to transmission in the BSs and the computing-plus-communication activities in the MEC server and $(F2)$ a quadratic term $(\xi(t) - L_{\text{in}}(t))^2$, which accounts for the QoS. In this regard, it is worth noting that F1 tends to push the system towards self-sustainability solutions and F2 favors solutions where the delay-sensitive load is entirely admitted in the MEC server by the router application, taking into account the expected energy to be harvested in the computing site. A weight $\Gamma = [0, 1]$ is utilized to balance the two objectives F1 and F2. The corresponding (weighted) cost function is defined as:

$$J(\delta, \psi, M, t) \triangleq \bar{\Gamma}\theta_{\text{EDGE}}(\delta_n(t), \{\psi_c(t)\}, M(t), t) + \Gamma(\xi(t) - L_{\text{in}}(t))^2, \quad (11)$$

where $\bar{\Gamma} = 1 - \Gamma$. Hence, starting from $t = 1$ (i.e., $t = 1, 2, \dots, T$) as the current time slot and the finite horizon T , the following optimization problem is formulated as:

$$P1 : \min_{\mathcal{E}} \sum_{t=1}^T J(\delta, \psi, M, t), \quad (12)$$

subject to:

$$A1: \delta_n(t) \in \{\varepsilon, 1\},$$

$$A2: \beta \leq C(t) \leq C,$$

$$A3: b(t) \geq b_{\text{low}},$$

$$A4: 0 \leq f_c(t) \leq f_{\text{max}},$$

$$A5: 0 \leq \lambda_c(t) \leq \lambda_{\text{max}},$$

$$A6: \chi_c(t) \leq \Delta,$$

$$A7: \sum_{c=1}^{C(t)} r_c(t) \leq r_{\text{max}},$$

$$A8: \theta_{\text{MEC}}(t) \leq b(t),$$

$$A9: \max \{2\rho_c(t)\} + \Delta = \tau_{\text{max}}, t = 1, \dots, T,$$

where the set of objective variables to be configured at slot t in the BS system and MEC server is defined as

$\mathcal{E} = \Delta\{\{\delta_n(t)\}, C(t), \{\psi_c(t)\}, \{P_c(t)\}, \{\lambda_c(t)\}, \zeta(t), M(t)\}$. The setting handles the transmission and computing-plus-communication activities. Constraint A1 specifies the BS operation status (either *power saving* or *active*), and A2 forces the required number of containers, $C(t)$, to be always greater than or equal to a minimum number $\beta \geq 1$: the purpose of this is to be always able to handle mission-critical communications. A3 makes sure that the EB level is always above or equal to a preset threshold β_{low} , to guarantee *energy self-sustainability* over time. Furthermore, A4 and A5 bound the maximum processing rate and workloads of each running container c , with $c = 1, \dots, C(t)$, respectively. Constraint A6 represents a hard-limit on the corresponding per-slot and per-VM processing time. A7 bounds the aggregate communication rate sustainable by the VLAN to r_{max} , and A8 ensures that the energy consumption at the computing site (due to the admitted computational workload) is bounded by the available energy in the EB. A9 forces the server to process the offloaded tasks within the set value τ_{max} .

From the optimization problem P1, it could be noted that $J(\zeta, \psi, M, t)$ consists of a nonconvex component defined in Eq. (7), while the others are convex and nondecreasing. In this case, Eq. (7) can be converted into a convex function using the GP concept [44], by introducing alternative variables and approximations. In this case, fixed parameters and approximations are introduced, i.e., μ_c, ν_c . In the sequel, the index t is dropped to improve readability. Thus, letting $r_c = 2\lambda_c/(\tau_{\text{max}} - \Delta)$ and then obtaining $P_c(r_c)$ in terms of λ_c by rearranging the Shannon-Hartley expression and substituting the value of r_c : $\hat{P}_c(r_c) = (((2\lambda_c/(\tau_{\text{max}} - \Delta)) - \nu_c W_c) \ln 2)/(\mu_c W_c) + \ln(N_0^{(c)}) - \ln g_c$. From the Shannon-Hartley expression, the presence of the log-sum-exp function is observed as it has been proven to be convex in [45] and recall that $P_c(r_c) = \exp(\hat{P}_c(r_c))$.

To solve P1 in (12), the LLC principles [28, 31], GP technique [44], and heuristics is used towards obtaining the feasible system control inputs $\varphi(t) = (\{\delta_n(t)\}, C(t), \{\psi_c(t)\}, \{P_c(t)\}, \{\lambda_c(t)\}, \zeta(t), M(t))$ for $t = 1, \dots, T$. Note that (12) can iteratively be solved at any time slot $t \geq 1$, by just redefining the time horizon as $t' = t, t + 1, \dots, t + T - 1$.

4.2. Edge System Management. In this subsection, a traffic load and energy harvesting prediction method and an online management algorithm are proposed to solve the previously stated problem P1.

4.2.1. Traffic Load and Energy Prediction. Given a time slot duration of $\tau = 30$, the time series prediction is performed, i.e., the $T = 3$ estimates of $\hat{L}_n(t)$ and $\hat{H}(t)$ are obtained by using an LSTM developed in Python using Keras deep learning libraries (Sequential, Dense) where the network has a one-dimensional (1D) subsequence of data, single feature, and multistep for an output. The dataset is split as 70% for training and 30% for testing. The efficient Adam implementation of stochastic gradient descent and fit the model for 20 epochs with a batch size of 4 is used. As for the performance measure of the model, the RMSE is used.

4.2.2. Edge System Dynamics. The system state vector at time t is denoted by $q(t) = (\delta(t), C(t), M(t), b(t))$, which contains the number of active BS, $\delta(t)$; number of active containers, $C(t)$; transmission drivers for data transfers, $M(t)$; and the EB level, $b(t)$. The input vector $\varphi(t) = (\{\delta_n(t)\}, C(t), \{\psi_c(t)\}, \{P_c(t)\}, \{\lambda_c(t)\}, \zeta(t), M(t))$ drives the MEC server behavior (handles the joint switching on/off of BSs, autoscaling and reconfiguration of containers, and the tuning of transmission drivers) at time t . In this work, $\{P_c^*(t)\}$ is obtained with CVXOPT toolbox, and $\{\lambda_c^*(t)\}$ is obtained by following the procedure outlined in Remark 1 (M. Andersen and J. Dahl. CVXOPT: Python Software for Convex Programming, 2019. [Online]. Available: <https://cvxopt.org/>).

The system behavior is described by the discrete-time state-space equation, adopting the LLC principles [14, 28]:

$$q(t+1) = \phi(q(t), \varphi(t)), \quad (13)$$

where $\phi(\cdot)$ is a behavioral model that captures the relationship between $(q(t), \varphi(t))$ and the next state $q(t+1)$. Note that this relationship accounts for the amount of energy drained $\theta_{\text{COMM}}(t)$, $\theta_{\text{MEC}}(t)$ that harvested $H(t)$ and that purchased from the electrical grid $E(t)$, which together lead to the next buffer level $\beta(t+1)$ through Eq. (10). The GENM algorithm finds the best control action vector that yields the desired energy savings within the edge network. Specifically, for each time slot t , problem (12) is solved, obtaining control actions for the prediction horizon T . The control action that is applied at time t is $\varphi^*(t)$, which is the first one in the retrieved control sequence. This control amounts to setting the number of active BSs, $\{\delta_n^*(t)\}$; number of instantiated containers, $C^*(t)$ (along with their obtained $\{\psi_c^*(t)\}, \{P_c^*(t)\}, \{\lambda_c^*(t)\}$ values); NIC status to either active or not, $\zeta^*(t) \in (0, 1)$; and the optimal transmission drivers, $M^*(t)$. The entire process is repeated every time slot t when the controller can adjust the behavior given the new state information.

State $q(t)$ and $\varphi(t)$ are, respectively, measured and applied at the beginning of the time slot t , whereas the offered load $L(t)$ and the harvested energy $H(t)$ are accumulated during the time slot and their value becomes known only by the end of it. This means that, being at the beginning of the time slot t , the system state at the next time slot $t+1$ can only be estimated, which is formally written as:

$$\hat{q}(t+1) = \phi(q(t), \varphi(t)). \quad (14)$$

Remark 1 (container provisioning and load distribution). For a fair provisioning of the computing resources, $C(t)$, and the expected workload allocation, $\hat{\xi}(t+1)$, a remark is presented. Firstly, each container can only compute an amount of up to λ_{max} , and to meet the latency requirements, $C(t)$ is obtained as $C(t) = \lceil (\hat{\xi}(t+1)/\lambda_{\text{max}}) \rceil$, where $\lceil \cdot \rceil$ returns the nearest upper integer. Secondly, to distribute the workload among the $C(t)$ containers, a heuristic process splits the computational workload $\lambda_c(t) = \lambda_{\text{max}}$ to the first $C(t) - 1$ containers, and the remaining workload $\lambda_c(t) = \hat{\xi}(t+1) - (C(t) - 1)\lambda_{\text{max}}$ to the last one.

4.2.3. Edge System Management Framework. In order to perform traffic load balancing using the green energy as a performance metric, a framework is defined that will identify the BS to be dynamically switched off and then steer the traffic load towards those BS with sufficient green energy. To do this, the available operating interval is defined as the ratio of the next time slot available green energy and the expected total power consumption (recall that the BS load is forecasted), per BS site, as

$$I_n(t) = \frac{b_n(t+1)}{\theta_{\text{BS},n}(t+1)} \geq 1. \quad (15)$$

If $I_n(t) < 1$, the BS site will not have sufficient energy to handle the expected traffic and it becomes a potential BS to be switched off. In the case where $I_n(t) \geq 1$, the site energy will be sufficient to handle the expected traffic. The potential BS to be switched off, denoted by BS n , will offload its traffic load to a neighboring BS, denoted by BS nn' . For BS nn' to be able to handle the offloaded traffic, the energy must be sufficient; thus, the green-based operating interval is defined as

$$x_{nn'}(t) = \frac{b_{nn'}(t+1)}{\theta_{\text{BS},nn'}(t+1)}, \quad (16)$$

where $\theta_{\text{BS},nn'}(t+1)$ is the total energy consumption of the BS site when the traffic load from the neighboring BS is combined with the expected load of the BS, and $b_{nn'}(t+1)$ is the next time slot energy. Next, the BS wake-up procedure is discussed.

(1) BS Wake-Up Procedure. To support BS reactivation commands, the UE location fingerprints that are obtained from the LS API are considered. The UE trajectory is assumed to be sequential, i.e., from BS-to-BS along the same direction (while still associated with the MEC server), and this is represented as $i_1(t) \rightarrow i_2(t) \rightarrow \dots \rightarrow i_n(t)$, where $i_n(t)$ refers to the serving/target BS node n in association with user i , at time slot t . When a BS node is switched off, it goes into a discontinuous reception cycle and configures a timer to awake and listen. Here, the MEC server, as BSs manager, sends wake-up control information as wake-up signaling (the information is a single bit). The wake-up information is only sent during the listening period. A BS n can be woken up only if it meets the following conditions: (i) $b_n(t+1) > b_{\text{low}}$ and (ii) a group of UEs that are associated with the MEC server is expected to receive the computed results via BS n (their trajectory is towards BS n as reported by the LS in the MEC).

4.2.4. Green-Based Edge Network Management (GENM) Algorithm. In order for the algorithm to manage the BS system, deciding upon the allocation of their transmission resources and also handling the computing and communication process, the best control action, $\varphi(t) = (\{\delta_n(t)\}, C(t), \{\psi_c(t)\}, \{P_c(t)\}, \{\lambda_c(t)\}, \zeta(t), M(t))$, that will yield the expected system behavior shall be obtained.

```

Input:  $q(t)$  (current state)
Output:  $\varphi^*(t)$  (control input vector)
01: Parameter initialization
    $\mathcal{S}(t) = \{q(t)\}$ .
02: for ( $k$  within the prediction horizon of depth  $T$ ) do
   -  $\hat{L}_{in}(t+k)$ := forecast the workload
   -  $\hat{H}_n(t+k)$ := forecast the energy
   -  $I_n(t+k)$ := operating interval of each BS
   -  $x_{nm}'(t+k)$ := green-based operating interval
   -  $\mathcal{S}(t+k) = \emptyset$ 
03: for (each  $q(t)$  in  $\mathcal{S}(t+k)$ ) do
   - generate all reachable states  $\hat{q}(t+k)$ 
   -  $\mathcal{S}(t+k) = \mathcal{S}(t+k) \cup \{\hat{q}(t+k)\}$ 
04: for (each  $\hat{q}(t+k)$  in  $\mathcal{S}(t+k)$ ) do
   - calculate the corresponding  $\theta_{EDGE}(\hat{q}(t+k))$ 
   taking into account of  $\kappa_n$  from  $L_{out}(t)$ 
   end for
   end for
05: - obtain a sequence of reachable states yielding
   minimum energy cost
06:  $\varphi^*(t) :=$  control leading from  $q(t)$  to  $\hat{q}_{min}$ 
07: Return  $\varphi^*(t)$ 

```

ALGORITHM 1. GENM algorithm pseudocode.

The edge network management algorithm pseudocode is outlined in Algorithm 1, and it is based on the LLC principles from [14, 28]. Starting from the *initial state*, the controller constructs, in a breadth-first fashion, a tree comprising all possible future states up to the prediction depth T . The algorithm proceeds as follows: a search set \mathcal{S} consisting of the current system state is initialized (line 01), and it is accumulated as the algorithm traverse through the tree (line 03), accounting for predictions, accumulated workloads at the output buffer, mobile devices trajectory $i_n(t)$, past outputs and controls, and operating intervals. The set of states reached at every prediction depth $t+k$ is referred to as $\mathcal{S}(t+k)$ (line 02). Given $q(t)$, the workload $\hat{L}_{in}(t+k)$ and harvested energy $\hat{H}(t+k)$ are estimated first, then obtain the operating intervals $I_n(t+k)$, $x_{nm}'(t+k)$ (line 02), and generate the next set of reachable control actions by applying the accepted workload $\xi(t+k)$, energy harvested, and green-based operating interval (line 03). The energy cost function corresponding to each generated state $\hat{q}(t+k)$ is then computed (line 04), where $\hat{q}(t+k)$ take into account of η_n as observed from $L_{out}(t)$. Once the prediction horizon is explored, a sequence of reachable states yielding minimum energy consumption is obtained (line 05). The control action $\varphi^*(t)$ corresponding to $\hat{q}(t+k)$ (the first state in this sequence) is provided as input to the system while the rest are discarded (line 06). The process is repeated at the beginning of each time slot t .

(1) *Algorithm Complexity.* The algorithm is executed at each time instance, and the corresponding time complexity is obtained as follows. The time complexity associated with the computation of the $I_n(t)$ and x_{nm}' is linear with the size

TABLE 2: System parameters.

Parameter	Value
Max. number of containers, C	20
Min. number of containers, β	1
Time slot duration, τ	30
Idle state energy for container c , $\theta_{idle,c}(t)$	4
Max. energy for container c , $\theta_{max,m}(t)$	10
Per-container reconfiguration cost, z_e	0.005 J/(MHz) ²
TOE in idle state, $\theta_{idle}^{NIC}(t)$	13.1 J
Max. allowed processing time, Δ	0.8
Processing rate set, $\{f_c(t)\}$	{0,50,70,90,105}
Bandwidth, W_c	1 MHz
Max. number of drivers, M	6
Noise spectral density, $N_0^{(c)}$	-174 dBm/Hz
Max. container c load, λ_{max}	10 MB
NIC best performance throughput, η	1.4 Gbit/J
Driver energy, $d_m(t)$	1 J/s
Target transmission rate, r_0	1 Mbps
Controllable factor of delay, u	0.96
Reconfiguration overhead, σ	20 ms
Leakage energy, $a(t)$	2 μ J
Energy storage capacity, b_{max}	490
Lower energy threshold, b_{low}	30% of b_{max}
Upper energy threshold, b_{up}	70% of b_{max}

of the BS group $|N|$ interconnected to the MEC server. Next, the complexity associated with updating the load allocation for the active BSs is $|N|-1$, which leads to $O(|N|^2)$. In the worst-case scenario (no BS has been switched off), the total complexity is $|N| |q(t)\varphi(t)T$, which is linear in all variables, namely, number of BSs interconnected to the MEC server, number of system states, number of control actions, and time horizon T .

5. Performance Evaluation

In this section, some selected numerical results for the scenario of Section 3 are shown. The parameters that were used in the simulations are listed in Table 2.

5.1. *Simulation Setup.* A virtualized MEC server in proximity to a group of BS is considered. The BS coverage areas overlap to enable load balancing. Our time slot duration τ is set to 30, and the time horizon is set to $T = 3$ time slots. For simulation, Python is used as the programming language.

5.2. *Numerical Results. Data preparation:* The information from the used mobile and energy traces is aggregated to the set time slot duration. The mobile traces are aggregated from 10 observation time to τ . As for the wind and solar traces,

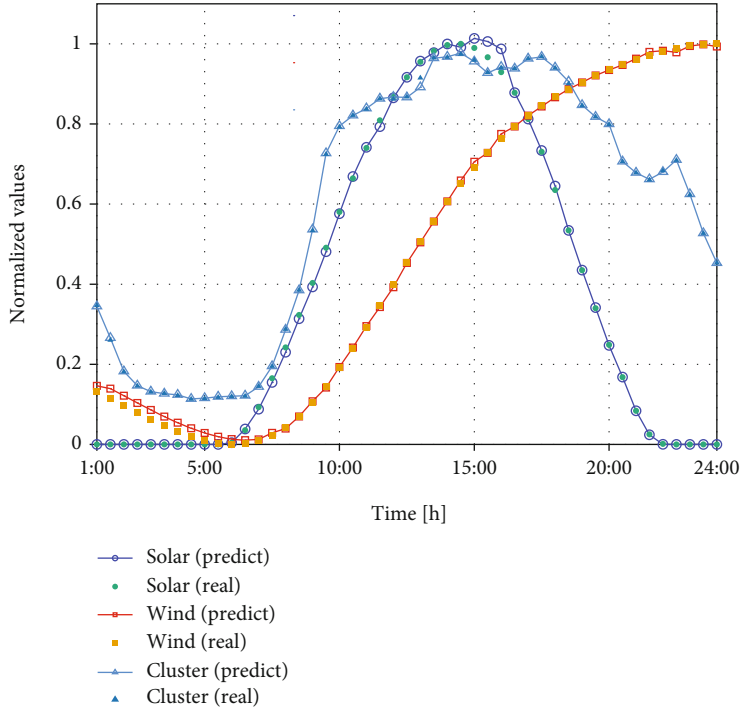


FIGURE 4: One-step ahead predictive values for $L(t)$ and $H(t)$.

they were aggregated from 15 observation time to τ . The used datasets are readily available in a public repository (see [39]).

In Figure 4, the real and predicted values for BS traffic load and harvested energy is shown. Here, the forecasting routing tracks each value and predicts it over one-step. The shown selected prediction results are for Cluster 3, Solar 3, and Wind 3. Then, Table 3 shows the average RMSE of the normalized harvested energy and traffic load processes, for different time horizon values, $T \in \{1, 2, 3\}$. In the table, the term $H_{\text{wind}}(t)$ represents the forecasted values for energy harvested from wind turbines, and $H_{\text{solar}}(t)$ is for the harvested energy from solar panels. From the obtained results, the prediction variations are observed between $H(t)$ and $L(t)$ when comparing the average RMSE. The measured accuracy is deemed good enough for the proposed optimization.

The GENM algorithm is benchmarked with another one, named Iterative-based Resource Manager with network impact Capability (IRMC), which is inspired by the iterative approach for computing platforms from [20] and the use of the network impact towards load balancing from [5]. Both algorithms make use of the learned information. Figures 5 and 6 show the average energy savings obtained by GENM in the MEC server. In Figure 5(a), the average results for GENM ($z_e = 0.005$, $|N| = 24$, $\Gamma = 0.5$, $\lambda_{\text{max}} = 10$ MB) show energy savings of 59%, while IRMC achieves 34% on average. As expected, the highest energy savings gain is observed in the early hours of the day (1-8) as the aggregated computational workload was at its lowest. In Figure 5(b), the average energy savings obtained by GENM is 68% ($z_e = 0.005$, $|N| = 12$, $\Gamma = 0.5$, $\lambda_{\text{max}} = 10$ MB) and for IRMC is 49%. Again, here, the highest peaks for energy savings are obtained from 1-8. The results are obtained with respect to the case where no

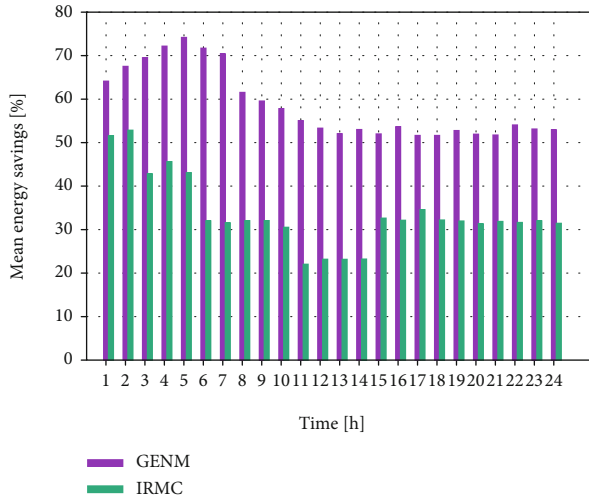
TABLE 3: Average prediction error (RMSE) for harvested energy and traffic load processes, both normalized in $[0,1]$.

	$T = 1$	$T = 2$	$T = 3$
$L(t)$	0.010	0.013	0.018
$H_{\text{wind}}(t)$	0.011	0.013	0.016
$H_{\text{solar}}(t)$	0.010	0.011	0.014

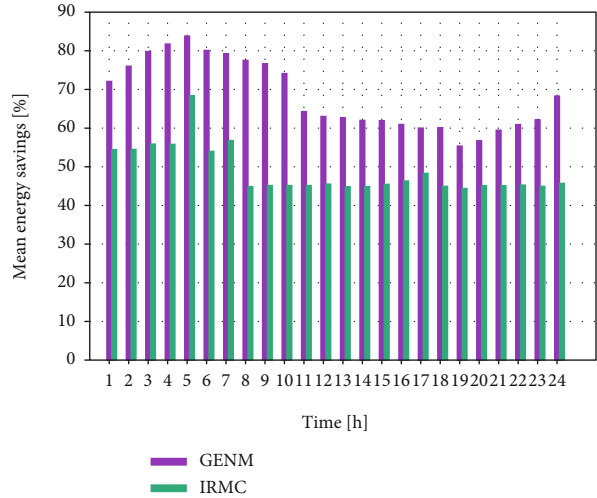
energy management procedures are applied, i.e., the MEC server provisions the computing resources for maximum expected computation workload (maximum value of $\theta_{\text{MEC}}(t)$, $C = 20$, $\forall t$). Comparing the results of Figures 5(a) and 5(b), we observed that when the BSs being manage by the MEC server are reduced (i.e., $12 < 24$), the aggregated delay-sensitive workload is also reduced and this translates to reduced computation process demands, which in turn results into high energy savings.

Figure 6 shows the average energy savings obtained when green energy is used as a performance metric towards load balancing within a group of BSs. Here, the group size is increased from $|N| = 5$ to 50, using an incremental step size of 5. The obtained energy savings are with respect to the case where all BSs are dimensioned for maximum expected capacity (maximum value of $\theta_{\text{COMM}}(t)$). From the results, it is observed that the energy savings increase as the BS cluster grows, thanks to the load balancing among active BSs.

Then, Figure 7 shows the average energy savings for the edge system. Here, the BS group size is set to $|N| = 40$, and the obtained energy saving results are with respect to the case where no energy management procedures are applied, i.e., the BSs are dimensioned for maximum expected capacity



(a) Mean energy savings for $\Gamma = 0.5, |N| = 24, \lambda_{\max} = 10$ MB



(b) Mean energy savings for $\Gamma = 0.5, |N| = 12, \lambda_{\max} = 10$ MB

FIGURE 5: Mean energy savings within the MEC server.

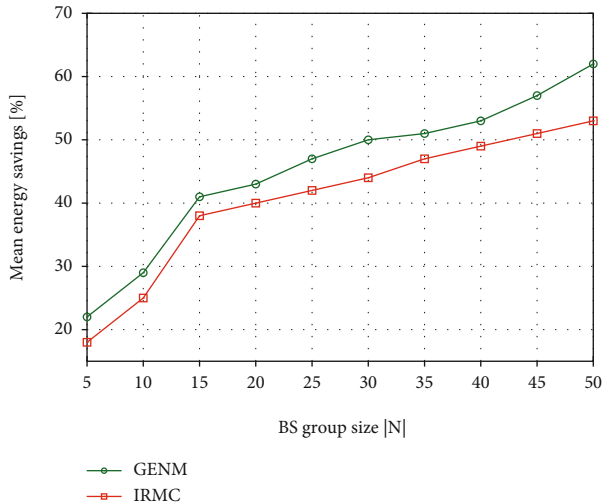


FIGURE 6: Energy savings versus BS group size.

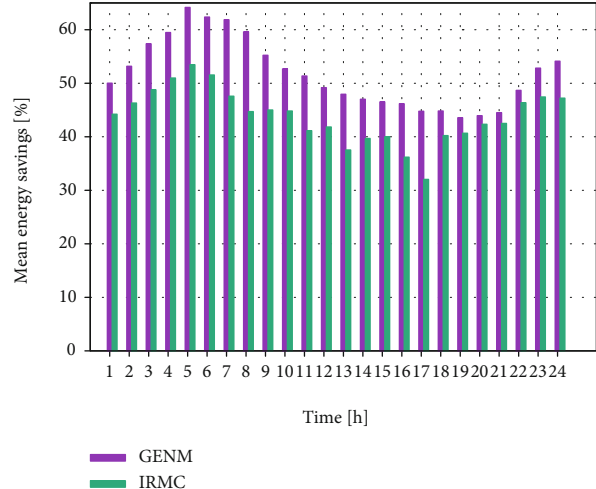


FIGURE 7: Mean energy savings for the edge system.

(maximum value of $\theta_{\text{COMM}}(t), \forall t$) and the MEC server provisions the computing resources for the maximum expected computation workload (maximum value of $\theta_{\text{MEC}}(t)$, with $C = 20$ containers, $\forall t$). The average results of GENM ($z_e = 0.05, \lambda_{\max} = 10$ MB, $\Gamma = 0.5$) show energy savings of 51%, while IRMC achieves 44% on average. The effectiveness of the joint dynamic BS management, autoscaling and reconfiguration of the computing resources, and on/off switching of the fast tunable laser drivers, coupled with foresighted optimization, is observed in the obtained numerical results.

6. Conclusions

This paper envisioned an edge system where a group of BS is placed in proximity to a MEC server for ease of handling the offloaded computational workload and BS management, and also, the edge apparatuses are powered by hybrid supplies, i.e., green energy is used in order to promote energy self-

sustainability within the network and as a performance metric for traffic load balancing. The extra energy can only be purchased from the grid supply to supplement the renewable energy supplies. The considered energy cost model takes into account the computing, caching, and communication processes within the MEC server, and the transmission-related energy consumption in BSs. To intelligently manage the edge system, an online algorithm based on forecasting, control theory, and heuristics is proposed with the goal of minimizing the overall energy consumption and to guarantee the quality of service within the network. The algorithm jointly performs (i) dynamic BS management using green energy as a performance metric; (ii) autoscaling and reconfiguration of the computing resources, workload, and processing rate allocation; and lastly, (iii) switching on/off of fast tunable drivers. Numerical results, obtained with real-world energy and traffic load traces, demonstrate that the proposed algorithm achieves energy savings of above 50% with respect to

the allocated maximum per-container loads of 10 MB. The computing platform is able to achieve energy savings from 59% to 68%, depending on the size of the BS cluster. The energy-saving results are obtained with respect to the case where no energy management techniques are applied in the BS system and the MEC server.

Data Availability

In this paper, open-source datasets for the mobile network (MN) traffic load and solar and wind energy have been used. The details are as follows: (1) the real MN traffic load traces used to support the findings of this study were obtained from the Big Data Challenge organized by Telecom Italia Mobile (TIM) and the data repository has been cited in this article. (2) The real solar and wind traces used to support the findings of this study have also been cited in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Morabito, V. Cozzolino, A. Y. Ding, N. Bejjar, and J. Ott, "Consolidate IoT edge computing with lightweight virtualization," *IEEE Network*, vol. 32, no. 1, pp. 102–111, 2018.
- [2] "Software-Defined and Cloud-Native Foundations for 5G Networks," https://www.interdigital.com/all_white_papers.
- [3] T. Han and N. Ansari, "A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 1038–1051, 2016.
- [4] J. Xu, H. Wu, L. Chen, C. Shen, and W. Wen, *Online Geographical Load Balancing for Mobile Edge Computing with Energy Harvesting*, IEEE International Conference on Communications (ICC), Kansas, USA, 2018.
- [5] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2126–2136, 2013.
- [6] T. Dlamini, A. F. Gambin, D. Munaretto, and M. Rossi, "Online supervisory control and resource management for energy harvesting BS sites empowered with computation capabilities," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 8593808, 17 pages, 2019.
- [7] R. Morabito, "Power Consumption of Virtualization Technologies: An Empirical Investigation," in *IEEE International Conference on Utility and Cloud Computing (UCC)*, pp. 522–527, Limassol, Cyprus, 2015.
- [8] Y. Jin, Y. Wen, and Q. Chen, "Energy Efficiency and Server Virtualization in Data Centers: An Empirical Investigation," in *2012 Proceedings IEEE INFOCOM Workshops*, pp. 133–138, Orlando, USA, 2012.
- [9] T. Dlamini, "Softwarization in future mobile networks and energy efficient networks," *Mobile Computing*, 2019, <https://www.intechopen.com/online-first/softwarization-in-future-mobile-networks-and-energy-efficient-networks=0pt>.
- [10] S. Fu, H. Wen, J. Wu, and B. Wu, "Cross-networks energy efficiency tradeoff: from wired networks to wireless networks," *IEEE Access*, vol. 5, pp. 15–26, 2017.
- [11] T. Dlamini and A. F. Gambin, "Adaptive resource management for a virtualized computing platform within edge computing," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1–9, Boston, USA, 2019.
- [12] M. Portnoy, *Virtualization essentials*, John Wiley and Sons, 2012.
- [13] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.
- [14] S. Abdelwahed, N. Kandasamy, and S. Neema, "Online Control for Self-Management in Computing Systems," in *Proceedings. RTAS 2004. 10th IEEE Real-Time and Embedded Technology and Applications Symposium, 2004*, pp. 368–375, Ontario, Canada, 2004.
- [15] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multiobjective auction-based switching-off scheme in heterogeneous networks: to bid or not to bid?," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9168–9180, 2016.
- [16] T. Han and N. Ansari, "On optimizing green energy utilization for cellular networks with hybrid energy supplies," *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 3872–3882, 2013.
- [17] L. Chen, S. Zhou, and J. Xu, "Computation peer offloading for energy-constrained mobile edge computing in small-cell networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1619–1632, 2018.
- [18] X. Jie and R. Shaolei, "Online learning for offloading and auto-scaling in renewable-powered mobile edge computing," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Washington, USA, 2016.
- [19] T. Dlamini, A. F. Gambin, D. Munaretto, and M. Rossi, "Online resource management in energy harvesting BS sites through prediction and soft-scaling of computing resources," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1820–1826, Bologna, Italy, 2018.
- [20] M. Shojafar, N. Cordeschi, D. Amendola, and E. Baccarelli, "Energy-saving adaptive computing and traffic engineering for real-time-service data centers," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1800–1806, London, UK, 2015.
- [21] M. Shojafar, N. Cordeschi, and E. Baccarelli, "Energy-efficient adaptive resource management for real-time vehicular cloud services," *IEEE Transactions on Cloud Computing*, vol. 7, no. 1, pp. 196–209, 2019.
- [22] M. Mukherjee, V. Kumar, S. Kumar et al., "Computation offloading strategy in heterogeneous fog computing with energy and delay constraints," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–5, Dublin, Ireland, 2020.
- [23] J. Bushra, S. Mohammad, A. Israr, U. Atta, M. Kashif, and I. Humaira, "A job scheduling algorithm for delay and performance optimization in fog computing," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 7, 2020.
- [24] M. Mithun, K. Suman, S. Mohammad, Z. Qi, and X. Mavromoustakis Constandinos, "Joint task offloading and resource allocation for delay-sensitive fog networks," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–7, Shanghai, China, 2019.

- [25] T. Zhao, S. Zhou, X. Guo, and Z. Niu, "Tasks scheduling and resource allocation in heterogeneous cloud for delay-bounded mobile edge computing," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–7, Paris, France, 2017.
- [26] B. Wu, S. Fu, X. Jiang, and H. Wen, "Joint scheduling and routing for QoS guaranteed packet transmission in energy efficient reconfigurable WDM mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 8, pp. 1533–1541, 2014.
- [27] D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang, "Power and performance management of virtualized computing environments via lookahead control," in *2008 International Conference on Autonomic Computing*, pp. 3–12, Chicago, USA, 2008.
- [28] N. Kandasamy, S. Abdelwahed, and J. P. Hayes, "Self-optimization in computer systems via on-line control: application to power management," in *International Conference on Autonomic Computing, 2004. Proceedings*, pp. 54–61, Washington, USA, 2004.
- [29] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, Melbourne, Australia, 2013.
- [30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [31] S.-L. Chung, S. Lafortune, and F. Lin, "Limited lookahead policies in supervisory control of discrete event systems," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1921–1935, 1992.
- [32] A. Ferdowsi, U. Challita, and W. Saad, "Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems: An Overview," *IEEE Vehicular Technology Magazine*, vol. 14, pp. 62–70, 2019.
- [33] J. Kumar, R. Goomer, and A. K. Singh, "Long short term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters," *Procedia Computer Science*, vol. 125, pp. 676–682, 2018.
- [34] S. Kekki, W. Featherstone, Y. Fang et al., *MEC in 5G Networks*, ETSI white paper, Sophia-Antipolis, France, 2018.
- [35] S. Ripduman, R. Andrew, A. W. Moore, and M. Kieran, "Characterizing 10 Gbps network interface energy consumption," in *IEEE Local Computer Network Conference*, pp. 268–271, Colorado, USA, 2010.
- [36] "Open big data challenge," <https://dandelion.eu/datamine/open-big-data/>.
- [37] D. Pelleg and A. W. Moore, "X-means: extending k-means with efficient estimation of the number of clusters," in *Proceedings of the 17th International Conf. on Machine Learning*, San Francisco, USA, 2000.
- [38] L. Chen, S. Zhou, and J. Xu, "Energy efficient mobile edge computing in dense cellular networks," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, 2017.
- [39] T. Dlamini, "Mobile and energy datasets," <https://github.com/lihles/mobile-datasets=0pt>.
- [40] M. Cardoso, M. R. Korupolu, and A. Singh, "Shares and utilities based power consolidation in virtualized server environments," in *2009 IFIP/IEEE International Symposium on Integrated Network Management*, pp. 327–334, New York, USA, 2009.
- [41] N. Cordeschi, M. Shojafar, and E. Baccarelli, "Energy-saving self-configuring networked data centers," *Computer Networks*, vol. 57, no. 17, pp. 3479–3491, 2013.
- [42] F. B. Abdesslem and A. Lindgren, "Large scale characterisation of YouTube requests in a cellular network," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, pp. 1–9, Sydney, Australia, 2014.
- [43] "Total generation," <https://www.elia.be/en/grid-data/power-generation>.
- [44] W.-C. Ho, L.-P. Tung, T.-S. Chang, and K.-T. Feng, "Enhanced component carrier selection and power allocation in LTE-advanced downlink systems," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 574–579, Shanghai, China, 2013.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

Research Article

Outage Probability Analysis of Decode-and-Forward Two-Way Relaying System with Energy Harvesting Relay

Hoang Van Toan,¹ Tran Manh Hoang ¹ and Le The Dung ^{2,3}

¹Telecommunications University, Khanh Hoa Province, Vietnam

²Division of Computational Physics, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

³Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Correspondence should be addressed to Le The Dung; lethedung@tdtu.edu.vn

Received 23 April 2020; Revised 29 October 2020; Accepted 22 November 2020; Published 7 December 2020

Academic Editor: Hideyuki Takahashi

Copyright © 2020 Hoang Van Toan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we evaluate a two-way relay system consisting of two terminals and an intermediate relay. In this model, two terminals do not have a direct link but exchange data with each other via the relay in three phases. The relay utilizes energy harvesting technology to collect energy from the received signals of two terminals in the first two phases and then uses the obtained energy for signal transmission in the third phase. Each node is equipped with a single antenna and operates under a half-duplex mode. All wireless channels are influenced by reversible independent flat Rayleigh fading. Using analytical methods, we provide the exact and approximate closed-form expressions of user outage probability and system outage probability. The approximate expressions of these outage probabilities are more explicit and straightforward, providing a better understanding of the influences of network parameters on the system quality. Monte-Carlo simulations are used to confirm the correctness of mathematical analyses.

1. Introduction

Nowadays, energy harvesting (EH) technology has become a hot research trend and attracted increasing interest from many research groups around the world [1–3]. It is a research trend towards green information in which network nodes can harvest the energy from different sources, such as solar, wind, vibration, thermoelectric effects, and other physical phenomena [4–6]. Based on the fact that radio frequency (RF) signals can carry energy and information at the same time [7], a new emerging solution for a wireless network is to avail ambient RF signals, in which wireless nodes can harvest energy and process the information simultaneously. Therefore, with EH technology, network nodes can collect energy from received signals in the radio band and use it for next operations [8, 9]. This operation not only helps to extend the operation life time of wireless nodes but also reduces their battery usage, resulting in a reduction of toxic wastage. Generally, based on the operation of receiver, EH technology is divided into two main techniques: (i) time

switching (TS), where the received signal is divided into two time phases to harvest energy and decode the signal, and (ii) power splitting (PS), where the received signal power is split into two parts for energy harvesting and signal decoding [9]. Based on receiver architecture, EH systems are divided into two main categories: (i) harvest-use, where all harvested energy is used immediately, and (ii) harvest-store-use, where harvested energy can be stored for the next operation [10]. For the receiver applying the TS scheme, its antenna is successively connected to the energy harvesting block and the signal processing block. The timing mechanism controls the signals coming to these blocks. On the contrary, for the receiver using the PS scheme, its antenna is connected to both the energy harvesting block and data processing block; thus, the received data are shared between them. Therefore, the EH phase and the information processing phase may concurrently occur in a given time slot, which shortens the transmission cycle. However, from the receiver's complexity perspective, the TS scheme is superior to the PS scheme because the commercially available circuits

separately designed for information decoding and energy harvesting can be used [11–13].

Recently, two-way relay systems have also been received much attention from researchers as they can improve spectral efficiency and extend the coverage area of wireless communication systems [14–17]. In [14], the authors study a system model where a bidirectional connection between two terminals is established via one amplify-and-forward (AF) or decode-and-forward (DF) relay and propose a novel relaying protocol where two relays alternately forward messages from a source terminal to a destination terminal, namely, two-path relaying. In [15], a two-way amplify-forward communication system employing the EH technique at relay node over Nakagami- m fading channels was considered. The authors gave the exact closed-form expression of user outage probability, the approximated expression of system outage probability, and the upper bound of ergodic capacity. The authors in [16] derived the expression of system outage probability and the sum-rate of a two-way network where two users exchanging information via multiple AF relays. To study the performance of a single relay selection scheme in a two-way relay system under outdated CSI conditions, the authors in [17] provided the exact and asymptotic expressions of the system outage probability. Similar to other systems, wireless nodes in two-way relay systems also encounter the problem of saving energy to extend the communication time. Therefore, the idea of combining EH technology in two-way systems is a new solution and has been studied in a number of recent publications. In [18], the authors considered a two-way relay cognitive radio system with a primary receiver. They evaluated this decode-and-forward (DF) relay system where the relay applies EH technology and network coding by deriving a closed-form expression of the system outage probability at terminals. However, this probability expression was not explicitly given. In [19], the authors studied the maximum throughput of a two-way relay system where two terminals and relay use EH technology. An energy harvesting two-way dual-relay network consisting of one non-EH relay and one EH relay with a finite-sized battery was investigated in [20]. The authors in [21] analyzed the outage probability and throughput of a three-phase two-way DF relay system model where the relay harvests the energy from radio frequency signals in the first two phases and converts it to transmission power in the third phase. However, in [21], the closed-form expression of overall system outage probability was not given. In [22], the authors considered three-step two-way decode-and-forward (DF) relay networks which are similar to the system model considered in [21]. Although they derived an analytical expression for the system outage probability and proposed a dynamic power splitting (PS) scheme to minimize the system outage probability, similar to [21], the exact expression of system outage probability was not given and the derivations of approximate expression are very complex and difficult to follow. The authors in [23] studied a two-way relaying system consisting of a multiantenna relay and two single-antenna sources. The relay was not properly powered but had to collect the energy from two sources' transmitted signals in the broadcast phase. Two sources exchanged

information in two time phases. The system performance in terms of outage probability and average throughput was analyzed. However, because the SNR expression was too complicated, they could not give closed-form expressions of these performance metrics. In [24], the authors consider relay beamforming and power-domain nonorthogonal multiple access for a wireless-powered multipair two-way relay network. Their objective was to optimize the energy transfer beamforming matrix by maximizing the minimum of the achievable rates among all the users. Applying a piecewise linear energy harvesting model for the users and the relay in dual-hop wireless powered two-way communication, the authors in [25] investigated the problem of total throughput maximization of both AF and DF relaying models. They concluded that it is essential to study a realistic EH model as the impractical linear EH model causes tremendous performance loss.

Motivated by the above issues, in this paper, we will derive the exact and approximate closed-form expression of overall system outage probability using the Taylor series expansion and Gaussian-Chebyshev quadrature approach. The approximation expressions are presented in a more concise form, allowing us to see the effects of key parameters such as the average transmission power, the channel gain, the time allocation of signal phase, and the power allocation coefficient on the system outage performance more easily. Particularly in this paper, the time allocation ratio of the signal phase and the power allocation coefficient are the distinguished parameters of the considered two-way DF relay system with EH relay; thus, it is necessary to clarify these parameters' influences on the outage performance of the considered system. Moreover, using the Taylor series expansion and Gaussian-Chebyshev quadrature approach allows us to adjust necessary precision by changing the number of terms in finite sum.

The rest of this paper is organized as follows. Section 2 outlines the proposed system model. The outage probability of this system model is studied in Section 3. Section 4 presents numerical results. Section 5 concludes the main findings of this paper.

2. System Model

In this paper, we consider a two-way relay system as shown in Figure 1. This system comprises two terminals S_1 and S_2 exchanging their data via an intermediate relay node R . The relay R is assumed to have a limited power and therefore has to harvest the energy from the radio frequency signals of two terminals to forward the information of these two terminals by using the DF three-phase two-way relay protocol. The motivation for using the DF relay are as follows: (i) the DF relay is found to be of more practical interest; (ii) compared with the AF relaying strategy, DF relaying avoids noise amplification and can be easily combined with coding technologies. All wireless channels are assumed to be reciprocal and undergo independent flat Rayleigh fading.

The three-phase two-way relay protocol can be described as follows. A cycle T for signal transmission between two terminals is divided into three phases. In the first phase t_1 , after

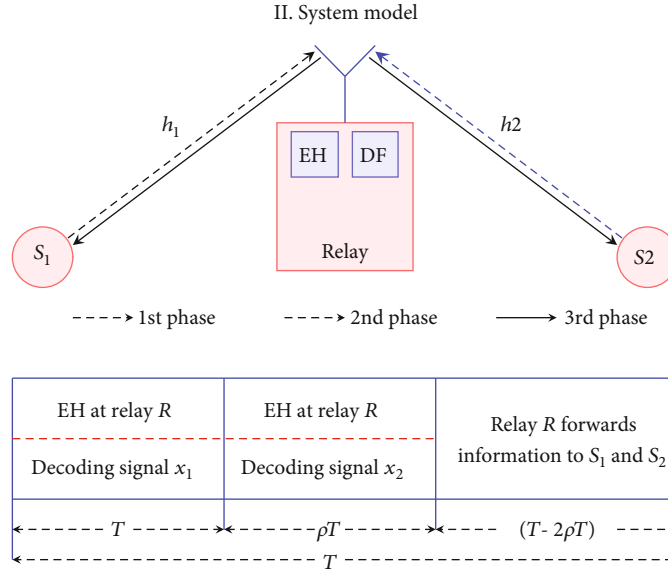


FIGURE 1: System model of the two-way DF relay system with EH relay.

R receives signals from S_1 , a power divider is used to divide the received power into two parts: one part is used to process signal and another part is used to convert energy. In the second phase t_2 , S_2 transmits its signal to R . Then, R divides the power of the received signal similarly to the first phase. It is assumed that $t_1 = t_2 = \rho T$, where $0 < \rho < 0.5$ is the time allocation ratio of the signal phase. In the third phase $t_3 = (1 - 2\rho)T$, the relay broadcasts the reencoding signals to two terminals.

We assume that the global channel state information (CSI) and partial CSI for the relay and terminals can be acquired perfectly. Specifically, for the relay R , the CSI of h_1 and h_2 have to be acquired to correctly decode the signal from two users simultaneously. For the terminal S_1 (S_2), the CSI of h_1 (h_2) have to be acquired to correctly decode the desired signals. In our considered relay system, the channel coefficients h_1 and h_2 can be obtained as follows. Terminal S_1 broadcasts a ready-to-send (RTS) message before information transmission. After receiving the RTS message from S_1 , terminal S_2 replies with a clear-to-send (CTS) message. Then, the relay R can estimate the channel coefficients of both h_1 and h_2 by overhearing the RTS and CTS messages. Finally, terminals S_1 and S_2 are informed of the corresponding channel coefficients through the feedbacks from the relay R [26].

Denote h_i , $i \in \{1, 2\}$, as the channel coefficients between S_i and R . When S_i transmits with power P , the received signal at R in the i th phase is

$$y_R^i = \sqrt{P}h_i x_i + n_i, \quad (1)$$

where n_i is an additive white Gaussian noise at the relay in the i th phase time, x_i is the transmitted signal symbol from S_i , and $\mathbb{E}\{x_i^2\} = 1$.

In the first two phases, R uses the divider to divide the received signal power into two parts: $\sqrt{\delta}y_R^i$ to harvest energy

and $\sqrt{1 - \delta}y_R^i$ to signal processing, where δ is the power allocation coefficient, $0 < \delta < 1$. The received power at R from S_i is given by

$$E_i = \eta \delta P |h_i|^2 \rho T, \quad (2)$$

where η is the energy conversion efficiency, $0 < \eta < 1$.

Consequently, we have the total energy which the relay collects in two phases as

$$E_{\sum \boxtimes} = \eta \delta P \rho T (|h_1|^2 + |h_2|^2). \quad (3)$$

It is worth noticing that the receiver consumes a certain amount of energy for CSI acquisition and circuitry. However, in our considered relay system, we assume the power required for CSI acquisition and circuitry is not supplied by the harvester energy but from an independent battery [10]. Thus, when all energy harvested in two phases is used for the signal transmission power of R in the third phase, the transmission power of R in the third phase is

$$P_R = \eta \delta P \left(\frac{\rho}{1 - 2\rho} \right) (|h_1|^2 + |h_2|^2). \quad (4)$$

According to the DF relaying protocol, R receives signals from S_1 and S_2 in two phases and decodes the received signals y_R^1 and y_R^2 into symbols x_1 and x_2 , respectively. Then, R encodes these two decoded symbols by applying XOR operation and obtains the normalized symbol x_R , i.e., $x_1 \oplus x_2 \rightarrow x_R$, where $\mathbb{E}\{x_R^2\} = 1$ and \oplus is the bitwise XOR operation.

In the third phase, R broadcast x_R to S_1 and S_2 . Terminal S_i receives this symbol and decodes it by using XOR operation with its transmitted symbol. Consequently, the received

signal at S_i in the third phase is given by

$$y_{S_i} = \sqrt{P_R} h_i x_R + n_{S_i}, \quad (5)$$

where n_{S_i} is the AWGN at S_i .

Since the noise power at all nodes in the system is equal to N_0 , the signal-to-noise ratios (SNRs) at R in the first and second phases are, respectively, calculated as

$$\gamma_1 = \frac{P(1-\delta)|h_1|^2}{N_0}, \quad (6)$$

$$\gamma_2 = \frac{P(1-\delta)|h_2|^2}{N_0}. \quad (7)$$

In the third phase, the SNR of the received signal at S_i , $i \in \{1, 2\}$, can be calculated as

$$\gamma_{3i} = \frac{\eta\delta P(\rho/(1-2\rho))(|h_1|^2 + |h_2|^2)|h_i|^2}{N_0}. \quad (8)$$

3. Performance Analysis

3.1. User Outage Probability (UOP)

3.1.1. Exact Expression. In the three-phase two-way relay protocol, the user outage probability of terminal S_i is the probability that S_i cannot successfully decode the intended received signal from S_j . In other words, the user outage probability of S_i is the probability that the SNR of the received signal at S_i is less than a threshold γ_{th} , i.e.,

$$UOP_i = \Pr(\gamma_{ji} < \gamma_{th}), \quad (9)$$

where $\gamma_{ji} = \min(\gamma_j, \gamma_{3i})$ with $i, j \in \{1, 2\}$, $i \neq j$ and $\gamma_{th} = 2^{3\mathcal{R}} - 1$; \mathcal{R} is the desired data transfer rate of S_i .

We can rewrite (9) as

$$\begin{aligned} UOP_i &= \Pr(\min(\gamma_j, \gamma_{3i}) < \gamma_{th}) = 1 - \Pr(\min(\gamma_j, \gamma_{3i}) > \gamma_{th}) \\ &= 1 - \Pr(\gamma_j > \gamma_{th}, \gamma_{3i} > \gamma_{th}) = 1 - \Pr\left(\frac{P(1-\delta)|h_j|^2}{N_0} > \gamma_{th}, \eta\delta P \right. \\ &\quad \left. \cdot \left(\frac{\rho}{1-2\rho}\right) \frac{(|h_i|^2 + |h_j|^2)|h_i|^2}{N_0} > \gamma_{th}\right). \end{aligned} \quad (10)$$

For clarity, we set $X = |h_i|^2$, $Y = |h_j|^2$, $\phi = \eta\delta\rho/(1-2\rho)$. Then, X and Y are exponential distributed random variables with mean parameters λ_x and λ_y , respectively. Hence, the UOP of S_i is computed as

$$UOP_i = 1 - \Pr\left(Y > \frac{\gamma_{th}N_0}{P(1-\delta)}, Y > \frac{\gamma_{th}N_0}{\phi PX} - X\right). \quad (11)$$

More specifically, the user outage probability at the terminal node S_i is defined in the following Theorem 1.

Theorem 1. *The user outage probability of S_i in the considered two-way DF communication system with EH relay is given by*

$$\begin{aligned} UOP_i &= 1 - e^{-((x_0/\lambda_x) + (\gamma_{th}N_0/P(1-\delta))\lambda_y)} \\ &\quad - \frac{1}{\lambda_x} \sum_{n=0}^N \mathbb{I} \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}\right)^n \left(\frac{\gamma_{th}N_0}{\phi P\lambda_y}\right)^{n/2} x_0^{(n+2)/2} e \\ &\quad - (\gamma_{th}N_0/2\phi P\lambda_y x_0) W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{th}N_0}{\phi P\lambda_y x_0}\right), \end{aligned} \quad (12)$$

where $x_0 = -(\gamma_{th}N_0/2P(1-\delta)) + (\sqrt{\gamma_{th}N_0} \sqrt{4P(1-\delta)^2 + \gamma_{th}N_0\phi/2P(1-\delta)}\sqrt{\phi})$, $W_{\alpha,\beta}(\cdot)$ denotes the function Whittaker ([27], Eq. (9.220)), and N is the number of truncated terms in the series expansion.

Proof. See Appendix A.

3.1.2. Approximate Expression in High SNR Regime. Herein, we use the Gaussian-Chebyshev quadrature approach to calculate the integral of a given function over interval (a, b) as

$$\int_a^b f(x) dx \approx \frac{b-a}{2} \sum_{v=1}^V \mathbb{I} \omega_v \sqrt{1-\gamma_v^2} f(x_v), \quad (13)$$

where $x_v = ((b-a)/2)y_v + (b+a)/2$, $y_v = \cos(((2v-1)/2V)\pi)$, and $\omega_v = \pi/V$.

Then, the approximate value of I_{1a} in (A.5) is

$$\tilde{I}_{1a} \approx \frac{1}{\lambda_x} \sum_{n=0}^N \mathbb{I} \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}\right)^n \frac{x_0}{2} \sum_{l=1}^L \mathbb{I} \omega_L \sqrt{1-\gamma_l^2} x_l^n e^{-(\gamma_{th}N_0/\phi P\lambda_y x_l)}, \quad (14)$$

where $\gamma_l = \cos(((2l-1)/2L)\pi)$, $\omega_L = \pi/L$, $x_l = x_0/2(\gamma_l + 1)$, and L is a parameter that determines the tradeoff between complexity and accuracy.

Finally, the approximate expression of the UOP of S_i is determined in the following corollary.

Corollary 2. *The approximate expression of the UOP of S_i is given by*

$$\begin{aligned} UOP_i^{P \rightarrow \infty} &\approx 1 - e^{-((x_0/\lambda_x) + (N_0/P(1-\delta))\lambda_y)\gamma_{th}} \\ &\quad - \frac{1}{\lambda_x} \sum_{n=0}^N \mathbb{I} \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}\right)^n \frac{x_0}{2} \sum_{l=1}^L \mathbb{I} \omega_L \sqrt{1-\gamma_l^2} x_l^n e^{-(\gamma_{th}N_0/\phi P\lambda_y x_l)}. \end{aligned} \quad (15)$$

Thanks to a more straightforward form than in (12), we can easily see the impacts of system parameters on each terminal's outage probability.

3.2. System Outage Probability (SOP)

3.2.1. Exact Expression. The outage probability of a two-way relay system is the probability that the SNR in at least one phase is lower than a given threshold γ_{th} .

Denote $\gamma_e = \min(\gamma_{3i}, \gamma_{3j})$, the system outage probability is calculated as

$$\text{SOP} = 1 - \Pr(\gamma_i > \gamma_{th}, \gamma_j > \gamma_{th}, \gamma_e > \gamma_{th}) = 1 - I_2 - I_3, \quad (16)$$

where

$$I_2 = \Pr(\gamma_i > \gamma_{th}, \gamma_j > \gamma_{th}, \gamma_e > \gamma_{th}, |h_i|^2 > |h_j|^2), \quad (17)$$

$$I_3 = \Pr(\gamma_i > \gamma_{th}, \gamma_j > \gamma_{th}, \gamma_e > \gamma_{th}, |h_i|^2 > |h_j|^2). \quad (18)$$

Then, the system outage probability of the considered system is defined in the following Theorem 3.

Theorem 3. *The exact closed-form expression of the system outage probability of the considered two-way DF relay system with EH relay is given by*

$$\text{SOP} = 1 - \exp\left(-\frac{\gamma_{th} N_0}{P(1-\delta)} \left(\frac{1}{\lambda_x} + \frac{1}{\lambda_y}\right)\right) \text{when } \frac{P < 2\phi\gamma_{th} N_0}{(1-\delta)^2}, \quad (19)$$

$$\begin{aligned} \text{SOP} = & 1 - e^{-((1/\lambda_x) + (1/\lambda_y))\sqrt{\gamma_{th} N_0 / 2\phi P}} \\ & - \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}\right)^n \left(\frac{\gamma_{th} N_0}{\phi P \lambda_x}\right)^{n/2} \\ & \cdot \left(x_1^{(n+2)/2} e^{-(\gamma_{th} N_0 / 2\lambda_x \phi P x_1)} W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{th} N_0}{\lambda_y \phi P x_1}\right) \right. \\ & \left. - \left(\frac{\gamma_{th} N_0}{P(1-\delta)}\right)^{(n+2)/2} e^{-((1-\delta)/(2\lambda_x \phi))} W_{(-n-2)/2, (n+1)/2} \left(\frac{1-\delta}{\lambda_y \phi}\right) \right) \\ & - \frac{1}{\lambda_y} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_y} - \frac{1}{\lambda_x}\right)^n \left(\frac{\gamma_{th} N_0}{\phi P \lambda_y}\right)^{n/2} \\ & \cdot \left(x_1^{(n+2)/2} e^{-(\gamma_{th} N_0 / 2\lambda_y \phi P x_1)} W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{th} N_0}{\lambda_x \phi P x_1}\right) \right. \\ & \left. - \left(\frac{\gamma_{th} N_0}{P(1-\delta)}\right)^{n+2/2} e^{-((1-\delta)/2\lambda_y \phi)} W_{(-n-2)/2, (n+1)/2} \left(\frac{1-\delta}{\lambda_x \phi}\right) \right) \text{when } \frac{P > 2\phi\gamma_{th} N_0}{(1-\delta)^2}, \end{aligned} \quad (20)$$

where $x_1 = \sqrt{\gamma_{th} N_0 / (2\phi P)}$.

Proof. See Appendix B.

In summary, depending on the transmission power P , the SOP is determined by (19) when $P < 2\phi\gamma_{th} N_0 / (1-\delta)^2$ or by (20) when $P > 2\phi\gamma_{th} N_0 / (1-\delta)^2$.

3.2.2. Approximate Expression in High SNR Regime. Recalling the Gaussian-Chebyshev quadrature approach in (13), the

approximate value of I_{2a} in (B.7) is as

$$\tilde{I}_{2a} \approx \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}\right)^n \frac{1}{2} \left(x_1 - \frac{\gamma_{th} N_0}{P(1-\delta)}\right) \sum_{k=1}^K \mathbb{X} \omega_K \sqrt{1 - \gamma_k^2} x_k^n e^{-(\gamma_{th} N_0 / \phi P \lambda_x x_k)}, \quad (21)$$

where $x_k = (1/2)y_k(x_1 - (\gamma_{th} N_0 / P(1-\delta))) + (1/2)(x_1 + (\gamma_{th} N_0 / P(1-\delta)))$, $y_k = \cos(((2k-1)/2K)\pi)$, $\omega_K = \pi/K$ with K is a parameter that determines the tradeoff between complexity and accuracy.

Applying a similar method for I_3 , we obtain an approximate expression of the SOP in the following corollary.

Corollary 4. *The approximate expression of the SOP is given by*

$$\text{SOP} = 1 - \exp\left(-\frac{\gamma_{th} N_0}{P(1-\delta)} \left(\frac{1}{\lambda_x} + \frac{1}{\lambda_y}\right)\right) \text{when } \frac{P < 2\phi\gamma_{th} N_0}{(1-\delta)^2}, \quad (22)$$

$$\begin{aligned} \text{SOP} \approx & 1 - e^{-((1/\lambda_x) + (1/\lambda_y))\sqrt{\gamma_{th} N_0 / 2\phi P}} \\ & - \frac{1}{2} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}\right)^n \left(x_1 - \frac{\gamma_{th} N_0}{P(1-\delta)}\right) \sum_{k=1}^K \mathbb{X} \omega_K \sqrt{1 - \gamma_k^2} x_k^n \\ & \cdot \left(\frac{1}{\lambda_x} e^{-(\gamma_{th} N_0 / \phi P \lambda_x x_k)} + \frac{1}{\lambda_y} e^{-(\gamma_{th} N_0 / \phi P \lambda_y x_k)}\right) \text{when } \frac{P > 2\phi\gamma_{th} N_0}{(1-\delta)^2}. \end{aligned} \quad (23)$$

It is noted that higher K results in a smaller difference in the approximate SOP expression. However, the value of K cannot be arbitrarily large because of the computational complexity. Fortunately, for our considered system model, even a small value of K ensures that the approximate expression well fits the exact expression, as demonstrated in the next section.

Remark 5. In this paper, the two-way relay system with one relay node is considered. In the case of a two-way multiple relay system, we must first model the relay selection scheme mathematically to find the outage probability (OP) expression. Then, the OP expression will contain the components representing the relay selection algorithm. In the case of a two-way relay system where the relay is equipped with multiple antennas, the channels between the relay and two terminals are random matrices. Furthermore, in the third phase time, the relay may employ a beamforming or transmit antenna selection (TAS) technique to transmit its signals. Thus, the OP expression will contain the components representing the beamforming vector at the relay or TAS algorithm. To sum up, it is challenging to find the OP expressions in these two scenarios.

4. Numerical Results

In this section, we use the exact and approximation expressions of the UOP and SOP obtained in the previous section to evaluate the two-way DF system's outage performance with the EH relay. Various scenarios are carried out to reveal

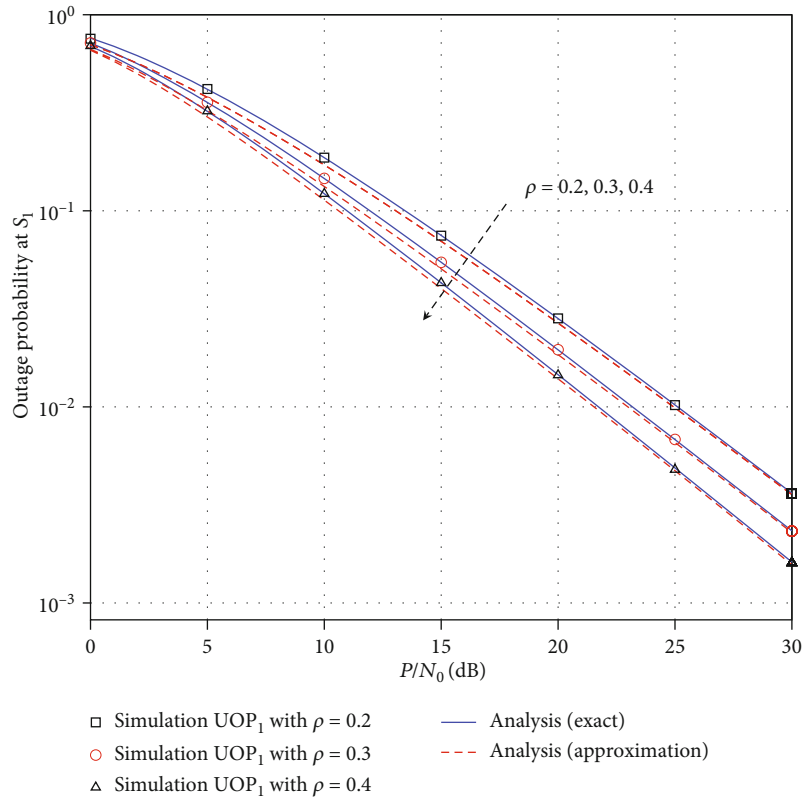


FIGURE 2: UOP of S_1 versus average SNRs for $\beta = 3$, $\gamma_{th} = 7$, $\eta = 0.8$, $\delta = 0.6$, $N_0 = 1$, $N = 5$, and $L = 5$.

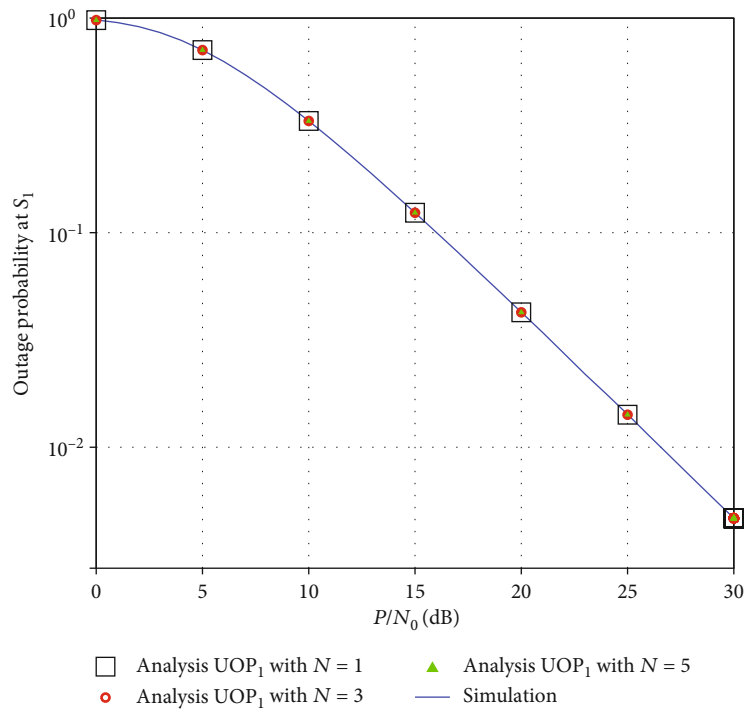


FIGURE 3: Effect of N on the UOP of S_1 for $\beta = 3$, $\gamma_{th} = 7$, $\eta = 0.8$, $\rho = 0.3$, $N_0 = 1$, and $L = 5$.

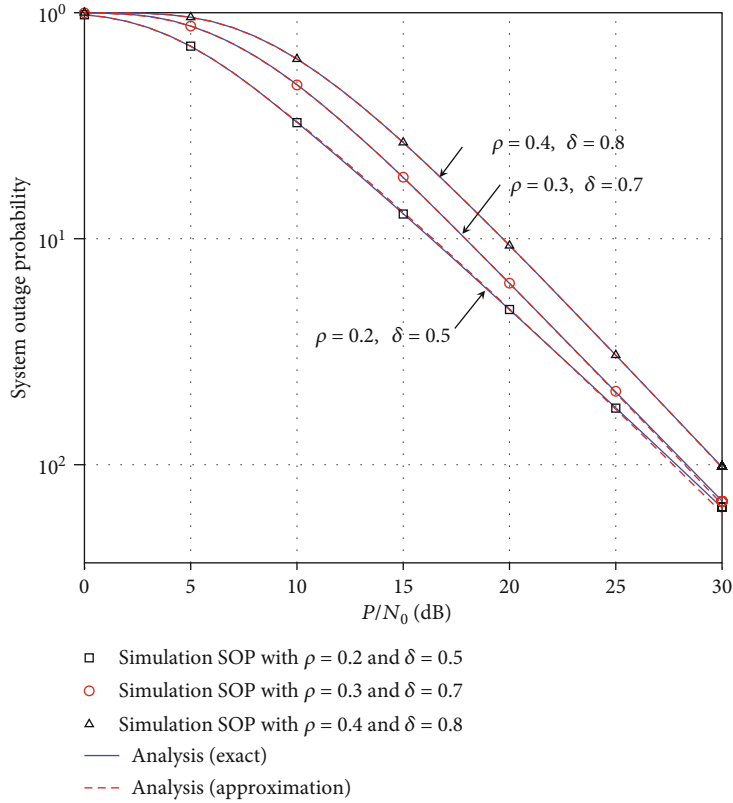


FIGURE 4: System outage probability versus average SNR for $\beta = 3, \gamma_{th} = 7, \eta = 0.8, N_0 = 1,$ and $N = K = 5$.

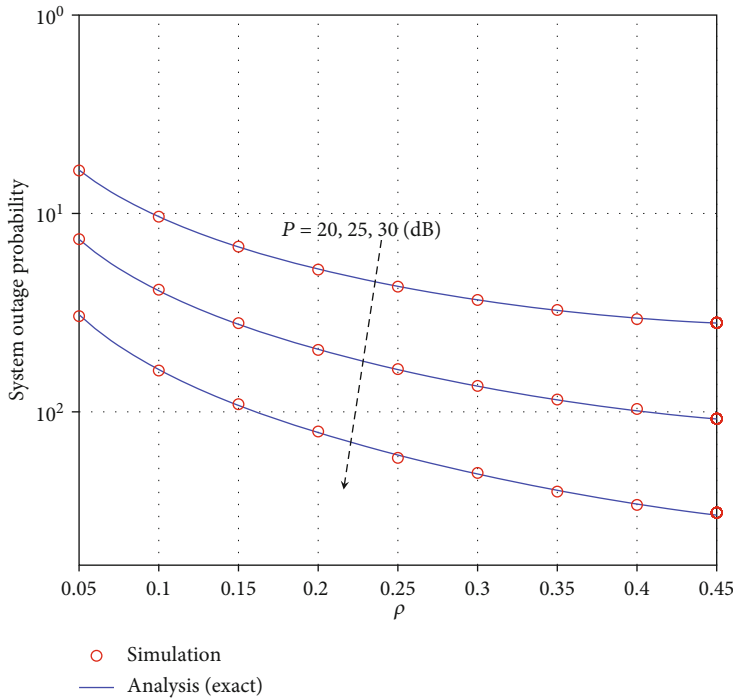


FIGURE 5: Effect of ρ on the system outage probability for different transmission power. $\beta = 3, \gamma_{th} = 7, \delta = 0.3, \eta = 0.8, N_0 = 1,$ and $N = K = 5$.

the impact of the Rayleigh fading, users' transmission power, the time allocation ratio, and the power allocation coefficient on the system performance. To demonstrate the accuracy of

the analysis results, we compare them with the Monte-Carlo simulation results. With regard to the simulation methodology, we first create the channel coefficients h_1 and

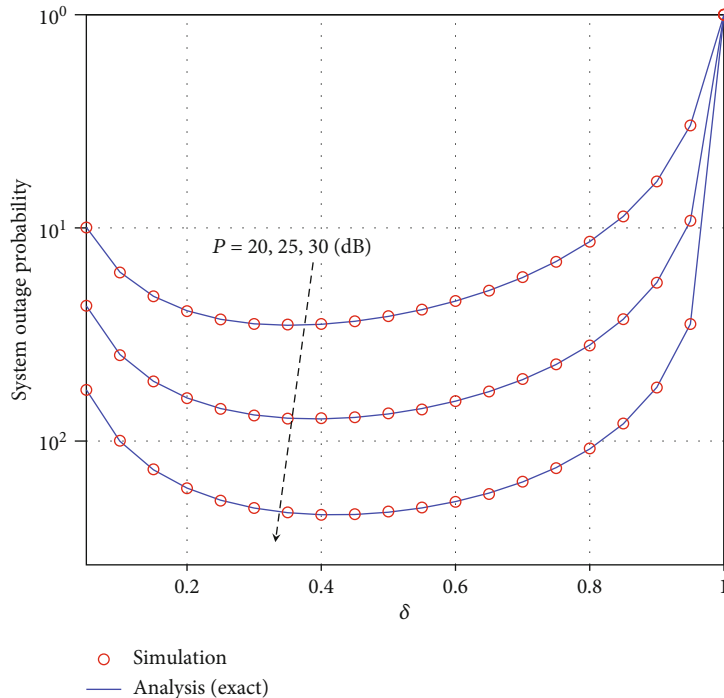


FIGURE 6: Effect of δ on the system outage probability for different transmission power. $\beta = 3$, $\gamma_{th} = 7$, $\eta = 0.8$, $\rho = 0.3$, $N_0 = 1$, and $N = K = 5$.

h_2 which are random variables with Rayleigh distribution with mean λ_x and λ_x , respectively. Then, we compute the SNR at the receivers. If this SNR is less than the predetermined threshold γ_{th} , then the outage happens. Such a process is repeated for a number of iterations of 10^6 . The outage probability is determined as the ratio of the number of times the outage events occur to the number of samples. It is assumed that network nodes are located on a 2D plane. The distance between the two terminals S_1 and S_2 is normalized to 1. Specifically, the locations of all nodes are $S_1(0,0)$, $S_2(1,0)$, and $R(0.4,0)$. We denote d_1 and d_2 as the physical distances from R to S_1 and from R to S_2 , respectively. For free-space path-loss transmission, we have $\lambda_x = d_1^{-\beta}$ and $\lambda_y = d_2^{-\beta}$, where β , $2 \leq \beta \leq 6$, is the path loss exponent. Unless otherwise stated, we set the following parameters $\beta = 3$, $\gamma_{th} = 7$, $\eta = 0.8$, $\delta = 0.6$, $N_0 = 1$, $N = 5$, and $K = L = 5$.

Figure 2 plots the exact and approximate analysis results and simulation results of the UOP of S_1 versus the average SNR P/N_0 for $\rho = 0.2, 0.3$, and 0.4 . We can see that the analysis results are in excellent agreement with the simulation ones, proving the correctness of the derived mathematical expressions. Moreover, the approximate results obtained by using the Gaussian-Chebyshev quadrature approach as in (15) are very close to the exact results. In addition, we observe that as ρ increases, the UOP of S_1 is reduced. It is because the increased ρ makes the transmission power in the third phase higher, thus improving the probability of successfully decoding signal at S_1 .

All theoretical analyses in this paper are based on the Taylor approximation method for exponential function such as in (A.4). For this approximation method, the accurateness can be enhanced by using more terms in the Taylor series. However, it may affect the computation time. To study the

accurateness matter, we conduct an evaluation of the UOP of S_1 for a different number of terms in Taylor series expansion, i.e., $N = 1$, $N = 3$, and $N = 5$, and give the results in Figure 3. We can see that a small number of terms in Taylor series expansion can provide a good match between the analysis results and simulation results. Particularly, even when $N = 1$, the analysis results are very similar to the simulation results. It means that Taylor series expansion is an efficient approximation method in this paper.

Figure 4 presents the system outage probability versus the average SNR for three cases: (1) $\rho = 0.2$, $\delta = 0.5$; (2) $\rho = 0.3$, $\delta = 0.7$; and (3) $\rho = 0.4$, $\delta = 0.8$. It can be seen from Figure 4 that the simulation results and analysis results are coincident, confirming the correctness of the analysis steps in this paper. We can observe that when increasing ρ and δ , the SOP decreases. It is because as ρ and δ gets higher, the harvested energy of R in two phases increases. Consequently, its transmission power in the third phase increases, resulting in higher SNR.

Figure 5 shows the SOP versus the time allocation ratio of signal phase ρ for three cases of the transmission power, i.e., $P = 20, 25$, and 30 dB. As observed in Figure 5, as ρ increases, the SOP decreases, indicating that the communication quality between S_1 and S_2 is better.

To further investigate the influence of the power-splitting ratio δ on the system performance, we evaluate the SOP as δ ranges from 0 to 1 and depict the result in Figure 6. We can see that based on the derived mathematical expression, the optimal δ at which the system outage probability is smallest can be determined. Specifically, the optimal value of δ in Figure 6 corresponding to $P = 20$ dB, $P = 25$ dB, and $P = 30$ dB are 0.34, 0.4, and 0.45, respectively.

5. Conclusion

In this paper, we have derived the exact and approximate closed-form expressions of the user outage probability and the system outage probability of a two-way DF relay system where the relay is not powered by a separated power supply but harvests the energy from the received radio frequency signals in two phases to convert it into the transmission power in the third phase. The approximate expressions of these two kinds of outage probabilities are in explicit and simplified forms, providing a better understanding of the effects of parameters on the quality of the system. Moreover, the accuracy of the approximate expressions can be adjusted by changing the number of terms in Taylor series expansion and the coefficients in the Gaussian-Chebyshev quadrature approach. All analysis results are verified by Monte-Carlo simulation results. Numerical results show that when the time allocation of signal phase ρ increases, the system outage probability decreases. Furthermore, using numerical results can determine the optimal power-splitting ratio δ at which the system outage probability is smallest. The exact and approximate closed-form expressions of the UOP and SOP in this paper provide a solid foundation for analyzing the performance of two-way relay systems with EH relay under other performance metrics such as ergodic capacity, energy efficiency, and average symbol error probability.

Appendix

A. Proof of Theorem 1

To support the derivation of the closed-form expression of UOP, we present two functions $y_1 = \gamma_{\text{th}} N_0 / P(1 - \delta)$ and $y_2 = (\gamma_{\text{th}} N_0 / \phi P x) - x$ in the Oxy coordinate plane to determine the integral domain as illustrated in Figure 7, where x_0 is the root of the equation $y_1 = y_2$.

Then, the UOP of S_1 can be calculated as

$$UOP_i = 1 - \underbrace{\int_0^{x_0} \int_{(\gamma_{\text{th}} N_0 / \phi P x) - x}^{\infty} f_{XY}(x, y) dy dx}_{I_{1a}} - \underbrace{\int_{x_0}^{\infty} \int_{(\gamma_{\text{th}} N_0 / P(1 - \delta))}^{\infty} f_{XY}(x, y) dy dx}_{I_{1b}}, \quad (\text{A.1})$$

where $f_{X,Y}(x, y)$ is the joint probability density function of two random variables X and Y , i.e.,

$$f_{XY}(x, y) = \frac{1}{\lambda_x} e^{-(x/\lambda_x)} \frac{1}{\lambda_y} e^{-(y/\lambda_y)}. \quad (\text{A.2})$$

First, I_{1a} can be computed as

$$I_{1a} = \int_0^{x_0} \int_{(\gamma_{\text{th}} N_0 / \phi P x) - x}^{\infty} \frac{1}{\lambda_x \lambda_y} e^{-(x/\lambda_x) + (y/\lambda_y)} dy dx \\ = \frac{1}{\lambda_x} \int_0^{x_0} \mathbb{E} e^{-((1/\lambda_x) - (1/\lambda_y))x - (\gamma_{\text{th}} N_0 / \phi P \lambda_y x)} dx. \quad (\text{A.3})$$

Applying Taylor series expansion for $e^{-((1/\lambda_x) - (1/\lambda_y))x}$, i.e.,

$$e^{-((1/\lambda_x) - (1/\lambda_y))x} = \sum_{n=0}^{\infty} \mathbb{E} \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n x^n, \quad (\text{A.4})$$

we have

$$I_{1a} = \frac{1}{\lambda_x} \sum_{n=0}^N \mathbb{E} \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \int_0^{x_0} \mathbb{E} x^n e^{-(\gamma_{\text{th}} N_0 / \phi P \lambda_y x)} dx, \quad (\text{A.5})$$

where N is the number of truncated terms in the series expansion.

With the help of [6] (Eq. (3.471.2)), we obtain

$$I_{1a} = \frac{1}{\lambda_x} \sum_{n=0}^N \mathbb{E} \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \left(\frac{\gamma_{\text{th}} N_0}{\phi P \lambda_y} \right)^{n/2} \\ \times x_0^{(n+2)/2} e^{-(\gamma_{\text{th}} N_0 / 2 \phi P \lambda_y x_0)} W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{\text{th}} N_0}{\phi P \lambda_y x_0} \right). \quad (\text{A.6})$$

Next, I_{1b} can be computed as

$$I_{1b} = \int_{x_0}^{\infty} \int_{\gamma_{\text{th}} N_0 / P(1 - \delta)}^{\infty} \frac{1}{\lambda_x \lambda_y} e^{-(x/\lambda_x) + (y/\lambda_y)} dy dx \\ = \int_{x_0}^{\infty} \frac{1}{\lambda_x} e^{-(x/\lambda_x)} \left(\int_{\gamma_{\text{th}} N_0 / P(1 - \delta)}^{\infty} \frac{1}{\lambda_y} e^{-(y/\lambda_y)} dy \right) dx \quad (\text{A.7}) \\ = e^{-((x_0/\lambda_x) + (\gamma_{\text{th}} N_0 / P(1 - \delta)) / \lambda_y)}.$$

Plugging (A.6) and (A.7) into (A.1), we get the closed-form expression of the UOP of S_1 as in (12).

B. Proof of Theorem 3

First, we find a closed-form expression of I_2 . Substituting (6), (7), and (8) into (17), we have

$$I_2 = \Pr \left(\begin{array}{l} X > \frac{\gamma_{\text{th}} N_0}{P(1 - \delta)}, Y > \frac{\gamma_{\text{th}} N_0}{P(1 - \delta)}, \\ (X + Y)X > \frac{\gamma_{\text{th}} N_0}{\phi P}, Y > X \end{array} \right) \\ = \Pr \left(X > \frac{\gamma_{\text{th}} N_0}{P(1 - \delta)}, (X + Y)X > \frac{\gamma_{\text{th}} N_0}{\phi P}, Y > X \right) \\ = \Pr \left(X > \frac{\gamma_{\text{th}} N_0}{P(1 - \delta)}, Y > \frac{\gamma_{\text{th}} N_0}{\phi P X} - X, Y > X \right). \quad (\text{B.1})$$

B.1. Case 1: The Transmission Power $P < 2\phi\gamma_{\text{th}}N_0/(1 - \delta)^2$. Considering the case when two terminals have transmission power satisfying the condition $P < 2\phi\gamma_{\text{th}}N_0/(1 - \delta)^2$, the

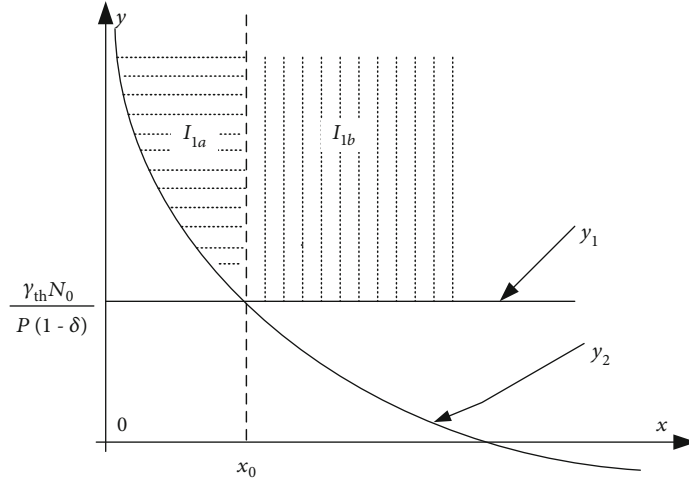


FIGURE 7: The integral calculation domain for I_{1a} and I_{1b} .

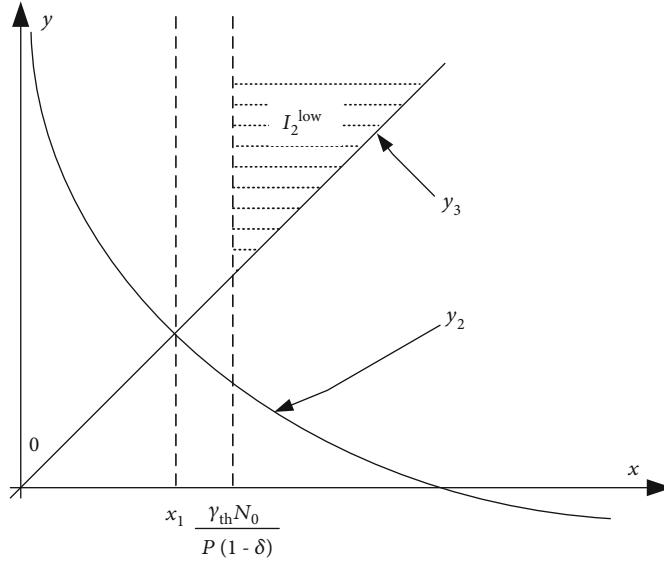


FIGURE 8: The integral calculation domain for I_2 in the low power region.

integral domain for I_2 in the low power region, I_2^{low} , is presented in Figure 8, where $y_2 = (\gamma_{\text{th}}N_0/\phi Px) - x$ and $y_3 = x$, and x_1 is the root of the equation $y_2 = y_3$.

Based on Figure 8, we have

$$I_2^{\text{low}} = \int_{\gamma_{\text{th}}N_0/P(1-\delta)}^{\infty} \int_{x_1}^{\infty} f_{X,Y}(x, y) dy dx, \quad (\text{B.2})$$

where $f_{X,Y}(x, y)$ is the joint probability density function of two random variables X and Y given in (A.2).

Substituting (A.2) into (B.2), we can calculate I_2^{low} as

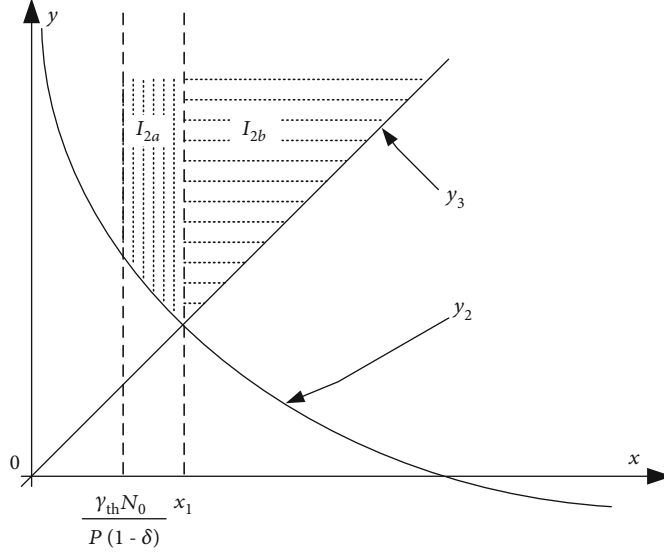
$$I_2^{\text{low}} = \frac{\lambda_y}{\lambda_x + \lambda_y} \exp\left(-\frac{\gamma_{\text{th}}N_0}{P(1-\delta)}\left(\frac{1}{\lambda_x} + \frac{1}{\lambda_y}\right)\right). \quad (\text{B.3})$$

Similarly, we can also calculate I_3 in the low transmission power region, I_3^{low} , as

$$I_3^{\text{low}} = \frac{\lambda_x}{\lambda_x + \lambda_y} \exp\left(-\frac{\gamma_{\text{th}}N_0}{P(1-\delta)}\left(\frac{1}{\lambda_x} + \frac{1}{\lambda_y}\right)\right). \quad (\text{B.4})$$

Plugging (B.3) and (B.2) into (16), we obtain the SOP in the low SNR region as in (19).

B.2. Case 2: The Transmission Power $P > 2\phi\gamma_{\text{th}}N_0/(1-\delta)^2$. Considering the case when two terminals have transmission power satisfying the condition $P > 2\phi\gamma_{\text{th}}N_0/(1-\delta)^2$, from (B1), we split the integral domain of I_2 in the high SNR region into two small regions I_{2a} and I_{2b} as in Figure 9.

FIGURE 9: The integral calculation domain for I_2 in the high power region.

Based on Figure 9, I_2 can be computed as

$$I_2 = \underbrace{\int_{\gamma_{th}N_0/P(1-\delta)}^{x_1} \int_{(\gamma_{th}N_0/P\phi x)-x}^{\infty} f_{XY}(x, y) dy dx}_{I_{2a}} + \underbrace{\int_{x_1}^{\infty} \int_x^{\infty} f_{XY}(x, y) dy dx}_{I_{2b}}. \quad (B.5)$$

The first part of (B.5), I_{2a} , can be calculated as

$$\begin{aligned} I_{2a} &= \int_{\gamma_{th}N_0/P(1-\delta)}^{x_1} \int_{(\gamma_{th}N_0/P\phi x)-x}^{\infty} \frac{1}{\lambda_x \lambda_y} e^{-(x/\lambda_x) - (y/\lambda_y)} dy dx \\ &= \int_{\gamma_{th}N_0/P(1-\delta)}^{x_1} \frac{1}{\lambda_x} e^{-(x/\lambda_x)} e^{-(1/\lambda_y)((\gamma_{th}N_0/P\phi x)-x)} dx \\ &= \frac{1}{\lambda_x} \int_{\gamma_{th}N_0/P(1-\delta)}^{x_1} \int e^{-x((1/\lambda_x) - (1/\lambda_y)) - (\gamma_{th}N_0/\lambda_y P\phi x)} dx. \end{aligned} \quad (B.6)$$

Recalling the Taylor series expansion in (A.4), we can rewrite (B.6) as

$$\begin{aligned} I_{2a} &= \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \int_{\gamma_{th}N_0/P(1-\delta)}^{x_1} \int x^n e^{-(\gamma_{th}N_0/\phi P \lambda_y x)} dx \\ &= \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \int_0^{x_1} \int x^n e^{-(\gamma_{th}N_0/\phi P \lambda_y x)} dx \\ &\quad - \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \int_0^{\gamma_{th}N_0/P(1-\delta)} \int x^n e^{-(\gamma_{th}N_0/\phi P \lambda_y x)} dx. \end{aligned} \quad (B.7)$$

Applying [27] (Eq. (3.471.2)), we obtain

$$I_{2a} = \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \left(\frac{\gamma_{th}N_0}{\phi P \lambda_y} \right)^{n/2} \cdot \left(\begin{array}{l} x_1^{(n+2)/2} e^{-(\gamma_{th}N_0/2\lambda_y \phi P x_1)} W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{th}N_0}{\lambda_y \phi P x_1} \right) \\ - \left(\frac{\gamma_{th}N_0}{P(1-\delta)} \right)^{(n+2)/2} e^{-(1-\delta)/2\lambda_y \phi} W_{(-n-2)/2, (n+1)/2} \left(\frac{(1-\delta)}{\lambda_y \phi} \right) \end{array} \right). \quad (B.8)$$

The second part of (B.5), I_{2b} , can be calculated as

$$\begin{aligned} I_{2b} &= \int_{x_1}^{\infty} \int_x^{\infty} f_{XY}(x, y) dy dx = \int_{x_1}^{\infty} \int_x^{\infty} \frac{1}{\lambda_x \lambda_y} e^{-(x/\lambda_x) - (y/\lambda_y)} dy dx \\ &= \frac{1}{\lambda_x} \int_{x_1}^{\infty} \int e^{-x((1/\lambda_x) + (1/\lambda_y))} dx = \frac{\lambda_y}{\lambda_x + \lambda_y} e^{-((1/\lambda_x) + (1/\lambda_y)) \sqrt{\gamma_{th}N_0/2\phi P}}. \end{aligned} \quad (B.9)$$

Substituting (B.8) and (B.9) into (B.5), I_2 finally becomes

$$\begin{aligned} I_2 &= \frac{\lambda_y}{\lambda_x + \lambda_y} e^{-((1/\lambda_x) + (1/\lambda_y)) \sqrt{\gamma_{th}N_0/2\phi P}} \\ &\quad + \frac{1}{\lambda_x} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y} \right)^n \left(\frac{\gamma_{th}N_0}{\phi P \lambda_y} \right)^{n/2} \\ &\quad \cdot \left(\begin{array}{l} x_1^{(n+2)/2} e^{-(\gamma_{th}N_0/2\lambda_y \phi P x_1)} W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{th}N_0}{\lambda_y \phi P x_1} \right) \\ - \left(\frac{\gamma_{th}N_0}{P(1-\delta)} \right)^{(n+2)/2} e^{-(1-\delta)/2\lambda_y \phi} W_{(-n-2)/2, (n+1)/2} \left(\frac{(1-\delta)}{\lambda_y \phi} \right) \end{array} \right). \end{aligned} \quad (B.10)$$

By using the above analysis steps for I_3 , the closed-form expression of I_3 is given by

$$I_3 = \frac{\lambda_x}{\lambda_y + \lambda_x} e^{-((1/\lambda_y) + (1/\lambda_x))\sqrt{\gamma_{th} N_0 / 2\phi P}} + \frac{1}{\lambda_y} \sum_{n=0}^N \frac{(-1)^n}{n!} \left(\frac{1}{\lambda_y} - \frac{1}{\lambda_x} \right)^n \left(\frac{\gamma_{th} N_0}{\phi P \lambda_x} \right)^{n/2} \cdot \left(\begin{array}{c} x_1^{(n+2)/2} e^{-(\gamma_{th} N_0 / 2\lambda_x \phi P x_1)} W_{(-n-2)/2, (n+1)/2} \left(\frac{\gamma_{th} N_0}{\lambda_x \phi P x_1} \right) \\ - \left(\frac{\gamma_{th} N_0}{P(1-\delta)} \right)^{(n+2)/2} e^{-((1-\delta)/(2\lambda_x \phi))} W_{(-n-2)/2, (n+1)/2} \left(\frac{1-\delta}{\lambda_x \phi} \right) \end{array} \right). \quad (\text{B.11})$$

Then, substituting (B.10) and (B.11) into (16), we obtain the closed-form expression of the SOP in the high SNR region as in (20).

Data Availability

Data are available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3622–3636, 2013.
- [2] N. Zlatanov, R. Schober, and Z. Hadzi-Velkov, "Asymptotically optimal power allocation for energy harvesting communication networks," *IEEE Transactions on Vehicular Technology*, vol. 99, pp. 1–1, 2017.
- [3] V. D. Nguyen, T. Q. Duong, H. D. Tuan, O. S. Shin, and H. V. Poor, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2220–2233, 2017.
- [4] K. Tutuncuoglu, B. Varan, and A. Yener, "Optimum transmission policies for energy harvesting two-way relay channels," in *2013 IEEE International Conference on Communications Workshops (ICC)*, pp. 586–590, Budapest, Hungary, 2013.
- [5] D. Wang, R. Zhang, X. Cheng, L. Yang, and C. Chen, "Relay selection in full-duplex energy-harvesting two-way relay networks," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 2, pp. 182–191, 2017.
- [6] H. Cao, L. Fu, and H. Dai, "Throughput analysis of the two-way relay system with network coding and energy harvesting," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, 2017.
- [7] L. R. Varshney, "Transporting information and energy simultaneously," in *2008 IEEE International Symposium on Information Theory*, pp. 1612–1616, Toronto, ON, Canada, 2008.
- [8] L. Liu, R. Zhang, and K. C. Chua, "Wireless information and power transfer: a dynamic power splitting approach," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3990–4001, 2013.
- [9] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: architecture design and rate-energy tradeoff," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, 2013.
- [10] I. Krikidis, Z. Gan, and B. Ottersten, "Harvest-use cooperative networks with half/full-duplex relaying," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 4256–4260, Shanghai, China, 2013.
- [11] Y. Gu and S. Aissa, "RF-based energy harvesting in decode-and-forward relaying systems: ergodic and outage capacities," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6425–6434, 2015.
- [12] B. Medepally and N. B. Mehta, "Voluntary energy harvesting relays and selection in cooperative wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3543–3553, 2010.
- [13] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information transfer with opportunistic energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 288–300, 2012.
- [14] R. Boris and W. Armin, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 379–389, 2007.
- [15] V. N. Q. Bao, H. Van Toan, and K. N. Le, "Performance of two-way AF relaying with energy harvesting over Nakagami-m fading channels," *IET Communications*, vol. 12, no. 20, pp. 2592–2599, 2018.
- [16] S. Atapattu, J. Yindi, J. Hai, and C. Tellambura, "Relay selection schemes and performance analysis approximations for two-way networks," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 987–998, 2013.
- [17] K. Hwang, M. Ju, and M. Alouini, "Outage performance of opportunistic two-way amplify-and-forward relaying with outdated channel state information," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3635–3643, 2013.
- [18] D. K. Nguyen, M. Matthaiou, T. Q. Duong, and H. Ochi, "RF energy harvesting two-way cognitive DF relaying with transceiver impairments," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1970–1975, London, UK, 2015.
- [19] K. Tutuncuoglu, B. Varan, and A. Yener, "Throughput maximization for two-way relay channels with energy harvesting nodes: the impact of relaying strategies," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2081–2093, 2015.
- [20] W. Li, M.-L. Ku, Y. Chen, K. R. Liu, and S. Zhu, "Performance analysis for two-way network-coded dual-relay networks with stochastic energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5747–5761, 2017.
- [21] N. T. P. Van, S. F. Hasan, X. Gui, S. Mukhopadhyay, and H. Tran, "Three-step two-way decode and forward relay with energy harvesting," *IEEE Communications Letters*, vol. 21, no. 4, pp. 857–860, 2017.
- [22] L. Shi, Y. Ye, R. Q. Hu, and H. Zhang, "System outage performance for three-step two-way energy harvesting DF relaying," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3600–3612, 2019.
- [23] H. Liang, C. Zhong, H. Lin, Y. Li, and Z. Zhang, "Optimization and analysis of wireless powered multi-antenna two-way relaying systems," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2048–2060, 2020.
- [24] C. Zhang and X. Jia, "Joint beamforming optimisation for NOMA-based wireless powered multi-pair two-way AF and

- DF relaying networks,” *IET Communications*, vol. 13, no. 4, pp. 387–395, 2019.
- [25] P. Ramezani and A. Jamalipour, “Two-way dual-hop WPCN with a practical energy harvesting model,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 8013–8017, 2020.
- [26] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, “A simple cooperative diversity method based on network path selection,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, 2006.
- [27] I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, and D. Zwillinger, *Table of Integrals, Series and Products, 7th ed.* Amsterdam, Elsevier, Boston, 2007.

Research Article

Low-Cost Authentication Protocol for D2D Communication in m-Health with Trust Evaluation

Ana Paula G. Lopes and Paulo R. L. Gondim 

Electrical Engineering Department, University of Brasilia, Brasilia, Distrito Federal, Brazil

Correspondence should be addressed to Paulo R. L. Gondim; prgond@gmail.com

Received 24 May 2020; Revised 9 August 2020; Accepted 16 September 2020; Published 27 October 2020

Academic Editor: Yujin Lim

Copyright © 2020 Ana Paula G. Lopes and Paulo R. L. Gondim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Relay-assisted device-to-device (D2D) communication serves users at the edge of system coverage of 5G networks, enabling communication among sensors and patients' mobile devices, and improving spectral and power efficiency. The security of D2D-based m-health applications requires attention due to the delicacy of the data treated in the collection, transmission, and storage of information on patients, whose devices must be adequately authenticated. However, traditional authentication and key agreement schemes are not suitable for D2D scenarios, since they might expose patients to security vulnerabilities and lead to an excessive use of resources. This article proposes a secure and lightweight scheme based on Shamir secret sharing for the mutual authentication of m-health devices in relay-assisted D2D communications, which provides security robustness and reduces resources (energy, processing) consumption. The manuscript also addresses the trustworthiness of devices involved in data relay and device discovery procedures.

1. Introduction

Mobile device communication has grown over the past few years due to the development of thousands of new applications and devices. The Internet of Things (IoT), the main responsible actor for such a revolution, enables the connection of several applications (e.g., those based on smartphones, smart watches, smart TVs, smart homes and vehicles, and smart metering). Mobile-health (m-health), which is an interesting human health-related application, provides the monitoring and evaluation of vital signs and other important health information on patients, preventing the escalation of diseases and affording immediate relief in emergencies.

The m-health system commonly works with a group of sensors coupled to a patient's body and a mobile device that receives the measurements from such sensors and sends the information to the respective health center. Huang et al. [1] observed high-quality healthcare services, such as remote monitoring, mobile telemedicine, remote disease diagnosis, and emergency care require the assurance of security of both

the system and the communication channels through which messages are exchanged.

On the other hand, D2D communication refers to direct and low-power communication between two mobile devices [1]; it offers services based on their proximity, and its advantages include higher throughput, low latency, and instantaneous communications between devices [2]. Moreover, traffic offloading/traffic steering between cellular and D2D networks is an excellent alternative for the bandwidth demands imposed over cellular networks, increasing spectral efficiency, and reducing energy consumption [1].

Device-to-device communication (D2D) is a strong candidate for communication of devices involved in m-health applications. For example, in a scenario of remote telemonitoring of patients implemented on a large scale by cellular and wireless body area networks (WBAN), the high volume of data exchanged, jointly with concurrent data traffic from other applications, requires a new perspective on the communication of near devices for providing important health information on patients' health, 24 hours a day and seven days a week. As another example, in emergency care

situations in which information about a patient must be rapidly transmitted for the evaluation and acting by a health professional (a doctor, for example), D2D communications contribute to the transmission and reception of data in a timely manner.

Health services provided at the edge of communication networks, such as LTE (Long-Term Evolution) and LTE-A (LTE-Advanced) can benefit from technological resources and principles inherited from mobile edge computing (or, more recently, multiaccess edge computing (MEC)) [3–6] discussed the integration of MEC and D2D technologies. In particular, a possible use of relay D2D devices can promote coverage extension and a better removal of constraints related to computation resources, since the task of a device can be off-loaded to an edge node and a nearby D2D device ([4]).

Nonetheless, Wang and Yan [2] highlighted the success of D2D communication depends on security, which has not been properly studied. D2D cannot work adequately to fulfill the application's expectations if security is not assured. Its requirements were addressed by Wang and Yan [2] and Haus et al. [7] and include authentication, privacy, anonymity, nonrepudiation, integrity, confidentiality, and resistance to attacks (e.g., man-in-the-middle, impersonation, and replay, among others).

Some of such security requirements might be fulfilled by a mutual authentication among devices and among devices and the core of the 3GPP (Third Generation Partnership Project) network. However, the traditional authentication and key agreement (AKA) standardized by 3GPP is not suitable for D2D authentication, since it leads to high consumption of processing and bandwidth resources as described by [8, 9]; moreover, D2D devices have commonly memory limitations and the energy (battery) consumption should also be reduced. Therefore, new applications that exchange critical data (e.g., m-health) require novel AKA schemes to fulfill such a demand [10].

This article proposes a novel mutual authentication and key agreement scheme (protocol) for D2D devices in m-health that enables patients to securely send their medical information to a health center and doctors. In comparison to other authentication protocols, our scheme reduces energy consumption and the use of processing resources.

An important issue regarding m-health is trust among devices supported by D2D communication. Whenever a patient must send data and no direct connection with the 3GPP infrastructure is provided, such data are sent through relay and device to device until the network infrastructure has been reached. The problem is that not all devices are trustworthy to perform such a task; consequently, trust assurance and evaluation become a critical problem for D2D m-health applications.

The protocol has been designed to forecast the relay of data when devices are outside the 3GPP coverage area, or inside it, but with no access to the network, regarding the necessity of computational trust.

The main contributions of the present study include the following:

- (i) A secure secret sharing scheme for D2D m-health applications that fulfills all security aspects discussed in 3GPP D2D security specification TS 33.303 [11]

- (ii) A mutual authentication scheme for D2D m-health groups of devices focused on the 3GPP architecture for proximity-based services
- (iii) An adaptation of the trust mechanism based on the local trust concept proposed by Yan et al. [12] that enables D2D devices to choose the most reliable device in their proximity to perform the relay of their data; in the proposed protocol, the local trust secret key encryption is based on symmetric cryptography, thus reducing computational costs when compared with [12], which is based on asymmetric cryptography
- (iv) An evaluation of computational and energy costs of our scheme, which revealed its superior performance, in comparison to two other protocols
- (v) An assessment of the security properties of the scheme and possible protection against attacks and threats
- (vi) A semiautomated formal validation of the protocol

The remainder of the manuscript is organized as follows: Section 2 discusses some related work. Section 3 presents the 3GPP reference architecture for proximity services. Section 4 introduces the protocol, its phases, the authentication and key agreement process, and a trust evaluation. Section 5 is devoted to a security analysis and comparisons with other protocols. Section 6 reports an evaluation of computational and energy costs. Section 7 discusses AVISPA verification. Section 8 provides the conclusions and suggests some future work.

2. Related Work

m-health security has been the focus of several studies. Zhang et al. [13] developed an efficient certificateless generalized signcryption (CLGSC) scheme based on the Elliptic Curve Discrete Logarithm Problem (ECDLP) and a lightweight and robust security-aware (LRSA) D2D-assisted data transmission protocol for m-health based on CLGSC. However, according to Zhou [14], the scheme proposed by [15] shows some security weaknesses, such as vulnerability to an insider attack, which affects its confidentiality. Zhou enhanced it by improving the CLGSC scheme and proposed a certificateless signcryption scheme for m-health [16], towards correcting the abovementioned vulnerabilities in CLGSC scheme. According to the author, it uses some extra variables in the authentication procedure in comparison to [15], which enables attackers to obtain some authentication parameters through queries.

Harn [15] presented three authentication schemes based on Shamir's secret sharing [17], which enables the generation of a common secret for a group of entities. According to Shamir's secret sharing, a previously established system manager chooses a random polynomial and generates a secret based on the secret tokens of each entity participating in the system. The tokens are then securely exchanged among the entities so that they reconstruct the secret through the Lagrange

interpolating formula and authenticate each other by comparing the secret generated with the secret received from the system manager.

Harn [15] designed the Asynchronous (t ; m ; n) Group Authentication Scheme (GAS) with Multiple Authentications, which authenticates n members of m groups and is resilient until t tokens have been compromised. Each entity has two tokens generated by the system manager through two different polynomials, which must remain secret. The system manager also generates a secret based on the tokens of the entities. Using its own two tokens, each member generates two Lagrange components, which are based on the Lagrange interpolating formula. The entities then exchange their Lagrange components to obtain a secret to be compared with that received from the system manager.

Mustafa and Philip [18] discussed the way a scheme of group key exchange for D2D medical IoT communication with cryptographic secret sharing must be designed to be efficient. Although it uses Shamir secret sharing [17], the authors do not detail the calculations and messages exchanged for the authentication of the devices, and only describe the procedure. A device is required to be a super-node that calculates the key generation process and distributes the key shares (tokens) to each device. The node is considered a single point of failure, since all devices rely upon it for the creation of the group-based session key. As future work, the authors propose the creation of a distributed key exchange approach. However, the development of a trust scheme for the D2D m-health environment has not been considered.

Yan et al. [12] designed a scheme for secure D2D communications that operates over the 3GPP infrastructure, based on two-dimensional trust levels, namely, Local Trust (LT), controlled by the communicating devices, and General Trust (GT), controlled by the 3GPP infrastructure. It considers D2D communication in general and presents the following three coverage scenarios: coverage, relay coverage, and out of coverage. The devices obtain support from ProSe Function Server (PFS) and ProSe App Server (PAS) to perform a trust evaluation. The scheme is composed of algorithms that authenticate and measure the trust level of devices in three situations, i.e., when only LT (local trust) level, or only GT (global trust), or both levels are used for the trust measurement, and has been partially used for the construction of our trust mechanism.

Last, but not least, we considered several technical reports and specifications of 3GPP regarding D2D communication and ProSe to strengthen the technical foundation of this study, including 3GPP TS 33.303 [11], 3GPP TS 23.303 [19], and TR 36.843 [20]. The former describes the security aspects to be considered when ProSe is used in the Evolved Packet System (EPS) and comprises the security procedures involving interfaces among network entities, the configuration of ProSe-enabled User Equipment (UE), and data transfer between ProSe Function and ProSe-enabled UE. The second specification [11] regards the ProSe features in EPS, i.e., ProSe discovery (identification of UEs in the proximity) and ProSe Direct Communication, which enables the establishment of communication paths between two or more

UEs in a direct communication range. The technical specifications in [20] address enhancements for ProSe UE-to-network relay for commercial and public safety applications, as wearables and IoT devices.

3. 3GPP Reference Architecture for ProSe Services

3.1. Functional Description. Since the development of applications and equipment can benefit from the adoption of standardized architectures, we considered the architecture adopted by a normative organization (in this case, the 3GPP) for proximity-based services.

The following entities of the 3GPP reference architecture for ProSe services have been considered:

- (i) Home Subscriber Server (HSS): part of the Evolved Packet Core (EPC) of LTE (Long-Term Evolution) networks that contains users' and subscribers' information, supports authentication and authorization of devices, and manages mobility
- (ii) ProSe Function Server (PFS): the logical function used for network-related actions required for ProSe that plays different roles for each of its feature [11], such as generation of trust tokens and identities in the management of D2D communication
- (iii) ProSe App Server (PAS): an entity that stores and manages ProSe User IDs and maintains permission information for restricted ProSe Direct Discovery
- (iv) User Equipment (UE): a mobile device associated with each user
- (v) Evolved NodeB (eNodeB): an entity that provides a wireless connection with UE and enables its connection with the core network

Figure 1 shows the reference architecture proposed by 3GPP for Proximity Services [20]. Domain A is inside the red dotted circle and comprises the security domain of EPC, PFS, and PAS. Domain B is defined by the lilac dotted circle and refers to the security domain of UE and PAS. Finally, Domain C defines the security domain comprised only of users' equipment.

3.2. Reference Points. Below is a list of reference points of 3GPP TS 23 303 [19], as shown in Figure 1:

- (i) PC1: the reference point between the ProSe application in the UE and the ProSe Application Server that defines application-level signaling requirements
- (ii) PC2: the reference point (PC2) between the ProSe Function Server (PFS) and the ProSe Application Server (PAS) that defines the interaction between PFS and PAS. PFS receives a proximity request from an originating UE and sends a proximity map request to PAS to obtain the identity of the

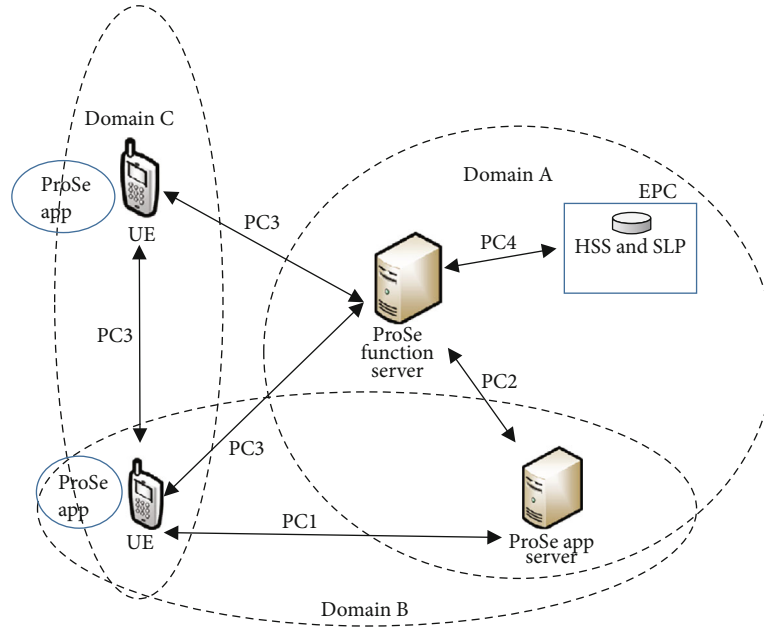


FIGURE 1: 3GPP ProSe reference architecture (based on Yan et al. [11]).

targeted application user. PAS determines whether the originating UE is allowed to discover the targeted UE

- (iii) PC3: the reference point between the UE and the ProSe Function that authorizes discovery requests in the EPC level and allocates the identities used in discovery procedures
- (iv) PC4: the reference point between HSS and PFS used by the latter to retrieve EPC-level discovery-related subscriber data
- (v) PC5: the reference point between UEs used for control and user plane for direct discovery

3.3. 3GPP ProSe Device Discovery. The 3GPP device discovery is detailed in technical specification TS 23.303 [19] and involves the detection and identification of other devices (UEs) located in proximities using E-UTRAN or WLAN direct radio signals. The device discovery can be open if no permission is required by the UE being discovered, or restricted, otherwise. It can also be used by applications to initiate ProSe Direct Communication.

It has the following two models for operations:

- (i) Model A (“I am here”): interested devices announce certain information in a predefined discovery interval, which could be used by devices nearby to obtain permission to discover their existence. They monitor the devices that showed interest in the messages, read, and process them
- (ii) Model B (“who is there?”/“are you there?”): devices transmit a request with the information on what they are interested in discovering. The addressed devices

respond with information related to the source device’s requests

Our scheme has adopted Model A of device discovery. First, each device must obtain authorization for direct discovery and direct communication from the PFS. Prior to announcing the information, they must send a discovery request to the PFS; if it succeeds, they can start announcing on the PC5 interface. Next, they send a request to the PFS to be authorized to monitor. If they succeed and have a Discovery Filter, they can start the monitoring. Finally, when the monitoring devices detect one or more devices that have matched the filter, they report them to the PFS. For a more detailed description, readers should consult 3GPP TS 23.303 [19].

3.4. Security Requirements. The several security requirements and aspects expected by the 3GPP standardization [19] for D2D communication that uses the ProSe architecture include the following:

- (i) Avoidance of attacks: the systems must resist several attacks, e.g., replay and impersonation
- (ii) Authorization of devices: the system must allow only currently authorized devices to be discovered by other UEs
- (iii) Tracking of devices: the tracking of devices based on their discovery messages should be minimized
- (iv) Authentication of devices and PFS: the devices involved must authenticate the source of the data received. UE and PFS must authenticate each other
- (v) Integrity and confidentiality: both integrity and confidentiality of data exchanged among the entities must be guaranteed

(vi) Privacy: the privacy of users must be provided

4. Proposed Scheme

Our scheme considers situations in which devices are outside the coverage area, in the coverage area, and directly connected with the 3GPP infrastructure, or in the coverage area, but with no direct access to 3GPP. In the second case, D2D nodes operate as the relay of a network, as proposed by Zhang et al. [13] and Zhou [14]. Moreover, computational trust is fundamental for a proper operation of the system.

HSS manages the device authentication and key distribution, whereas PFS and PAS manage the trust of devices. D2D communication involves patient's devices willing to perform the relay of data. Finally, the health center infrastructure receives the patients' data and forwards them to doctors, nurses, and physicians. Figure 2 shows the architecture of the protocol, derived from 3GPP ProSe [11] standards, with all entities involved.

Table 1 shows the main symbols and parameters used in the proposal.

Some basic assumptions are as follows:

- (i) The health center infrastructure is considered trustworthy and secure
- (ii) The entities of the 3GPP infrastructure and their communication channels are considered trustworthy and secure
- (iii) The channel between the patients' device and their respective body sensors is considered safe
- (iv) The D2D communication channels and the channel between devices and eNB are considered unsafe (they are the focus of this study)

The domain considered covers one or more 3GPP cells. Several groups can be inside the system domain of operation and are formed according to the patients' needs regarding the sending of their data.

The protocol uses asymmetric and symmetric cryptography. The former is used in the generation of private keys and temporary identities for mutual authentication, while symmetric cryptography is employed in the trust evaluation for reducing costs, when compared to [12]. It is based on the Asynchronous (t ; m ; n) Group Authentication Scheme (GAS) with Multiple Authentications, proposed by Harn [15], since it provides a way of sharing a secret in a group of entities that might be used in the generation of secret keys. Timestamps and random variables are freshly generated in each session for avoiding attacks. A session key is generated among devices, as well as among devices and HSS at the end of the mutual authentication phase, and a local trust key is generated whenever a local trust evaluation is required from one device to another. New keys are generated at every single execution of the protocol.

The following 5 (five) phases, described in the next subsections, must be executed for a patient outside the coverage area to send their data in a protected manner:

- (i) Initialization
- (ii) Registration
- (iii) Mutual authentication
- (iv) Trust evaluation
- (v) Encryption/decryption

4.1. Initialization. Some important system parameters are generated in this phase, and all devices accredited by the health center server must perform the phase offline.

Initially, HSS selects two random prime numbers p and q that satisfy condition $q|(p-1)$ and defines a finite field Zp^* and a secure elliptic curve $E(Zp^*)$. Next, it selects a group Gp of order p , Gq that is a subgroup of Gp , g as the generation point of Gq , and Zq^* as a prime field of order q . Then, it selects a random number $z \in Zq^*$ as the master private key and calculates $M_{kpub} = z * g$ to obtain the master public key.

HSS selects three hash functions ($h1(\cdot)$, $h2(\cdot)$, and $h3(\cdot)$) (described in Table 1) for the mutual authentication phase and generates j random numbers, R_j , ($j = 1, 2, \dots, i$), for each device and for itself for the calculation of a set of temporary identities TID_{D_i} :

$$TID_{D_i} = h_1(ID_{D_i} || R_j * z). \quad (1)$$

It also selects its own TID_{HSS} :

$$TID_{HSS} = h_1(ID_{HSS} || R_j * z). \quad (2)$$

Next, it sends each device its respective set. A different TID_{D_i} is used whenever a new session has been established to provide a relay of data to a specific device. When the last TID available is being used, the device must notify the HSS after the authentication procedure. Then, HSS sends a new set of temporary identities encrypted with the freshly generated session key.

HSS generates a piece of each device's private key (similar to [15]), chooses a $y \in Zq$, and calculates the following:

$$Y_{D_i} = h_2(TID_{D_i} || y) * M_{Kpub}. \quad (3)$$

Finally, it sets the partial key calculated for each respective device and publishes the following parameters: $\{g, G, E(Zp^*), M_{kpub}, TID_{HSS}, h1(\cdot), h2(\cdot), h3(\cdot), H(\cdot)\}$.

4.2. Registration. The ProSe device discovery mechanism is initially applied in this phase for the discovery of nearby devices, as described in [11]. The phase is performed over an insecure channel, and the main steps are described below.

Each user generates a share of its private key (based on [13]) choosing $x \in Zq^*$ and calculates its public key:

$$PK_{D_i} = x * Y_{D_i} * g. \quad (4)$$

Then, it sends TID_{D_i} , PK_{D_i} , and a timestamp T_{D_i-1} to the other devices and the nearby device sends HSS all the information received from relay devices.

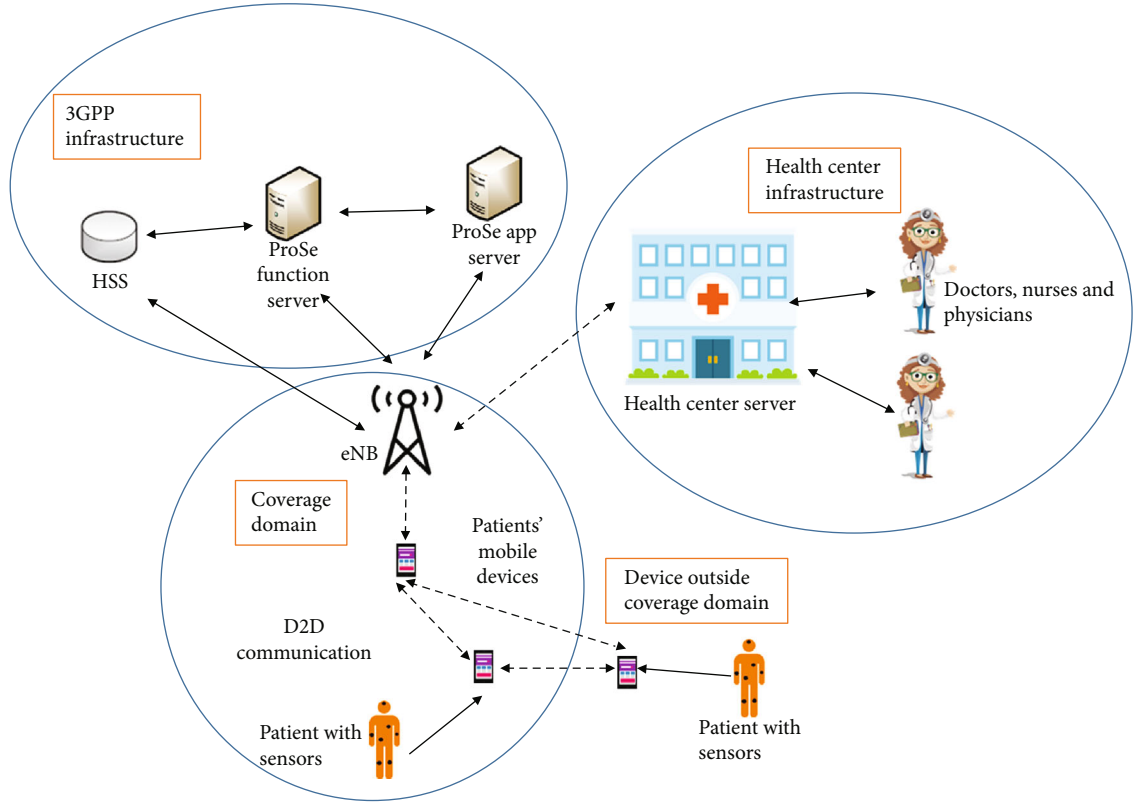


FIGURE 2: System architecture.

The device sets its private key as pair $SK_{D_i} = (x, Y_{D_i})$ using the other share of its private key received from the HSS in the initialization phase.

Each device chooses an integer $v_{D_i-D_j} \in Zq^*$ as a secret value to be sent to other devices and HSS, encrypted with its public key, as follows:

$$V_{D_i-D_j} = \text{EPK}_{D_i}(v_{D_i-D_j}), \quad (5)$$

where j represents either a device D_j or HSS.

Consequently, only the correct device can decrypt the message and obtain the secret token. The secret values are broadcast to the entities involved in the communication, which find and decrypt them to obtain all secret values necessary for the generation of session keys.

The asynchronous mode of the group authentication protocol designed by Harn [15] is considered for providing multiple authentications in a t -secure m -user n -group authentication scheme (GAS). In other words, for a group with n members, m -users are authenticated at once, with at most $(t - 1)$ compromised tokens; a unique token is assigned for each user of a group by the group manager for determining the membership of a user to a group. Therefore, considering what is proposed by [15], we have designed our authentication scheme as follows:

First, HSS selects two random polynomials $f_l(x), l = 1, 2$ of degree $t - 1$ each, where $t \leq n$ is the number of devices involved in the relay (i.e., number of tokens necessary for the recovery of secret S):

$$f_l(x) = \sum_{i=0}^{t-1} a_i x^i \text{ mod } p. \quad (6)$$

All coefficients are in finite field Zp^* .

HSS generates two tokens for each device calculating $f_l(\text{TID}_{D_i})$. Each TID_{D_i} has its respective token. HSS also calculates its own two tokens $f_l(\text{TID}_{\text{HSS}})$ and finds integers $w_{i,j}$, $d_{i,j}, j = 1, 2 \in Zp^*$, such that $S = \sum_{j=0}^k d_{i,j} f_j(w_{i,j})$, where $w_{i,1} \neq w_{i,2}$ for every pair i and j . It then generates a secret S , as in [17]:

$$S = f_l(0) = a_0, \quad (7)$$

$$S = g^{\sum_{j=1}^2 d_{i,j} f_j(w_{i,j}) \text{ mod } q} \text{ mod } p. \quad (8)$$

TABLE 1: Parameters used in the protocol.

Symbol	Description
D_i	Patient i or device i , where $i = 1, 2, 3, \dots, n$
p, q	Large public prime numbers
$Zp * /Zq *$	A finite field of order p /prime field of order q
$E(Zp)$	Elliptic curve over Zp
Gp	Group of order p
Gq	A subgroup of Gp of order q
g	Point generator of G ; $i = 1, 2, 3 \dots$
$f_l(x)$	Random polynomial; $l = 1, 2, \dots$
z	Master private key
M_{kpub}	Master public key
SK_{D_i}	Private key of device D_i
PK_{D_i}	Public key of device D_i
ID_x/TID_x	Real identity of entity x /temporary identity of entity x
R_j	j random number generated
$T_{x,i}$	Timestamp generated by entity $x = D_i, HSS$; $i = 1, 2, 3 \dots$
h_1	Temporary identity generation hash function; $H_1 : \{0, 1\} * Zp *$
h_2	Device's partial private key generation hash function; $H_2 : \{0, 1\} * Gq$
h_3	Symmetric key generation hash function; $H_4 : \{0, 1\} * \times GqGp$
H	Shamir's secret hash function
LT_i	Local trust level of device i
$LTK_{D_u-D_u'}$	Local trust secret key
	Secure channel
	Insecure channel

Finally, it chooses an integer $v_{HSS-D_i} \in Zq *$ and sends it to the respective devices of the relay group:

$$AUTH_{D_i} = [EPK_{D_i}(TID_{HSS}, H(S), w_{i,j}, d_{i,j}, f_l(TID_{D_i}), v_{HSS-D_i}), TID_{D_i}, T_{HSS-1}]. \quad (9)$$

The devices decrypt the message and store the parameters.

Figure 3 shows a summary of the registration phase.

4.3. Mutual Authentication. Since the devices still must authenticate each other and HSS, each device selects a pair of nonused TID_{D_i} and respective tokens $f_l(TID_{D_i})$, $l = 1, 2, \dots$ and computes its Lagrange component (an adaptation of what is proposed in [15]) LC_{D_i} through the Lagrange interpolating formula:

$$LC_{D_i} = \sum_{l=1}^2 d_{i,l} f_l(TID_{D_i}) \prod_{q=1; q \neq i}^n \frac{w_{i,l} - TID_{D_i-q}}{TID_{D_i} - TID_{D_i-q}} \mod q. \quad (10)$$

Next, they calculate $e_{D_i} = g^{LC_{D_i}} \mod p$.

HSS also calculates its Lagrange component LC_{HSS} through the Lagrange interpolating formula:

$$LC_{HSS} = \sum_{l=1}^2 d_{i,l} f_l(TID_{HSS}) \prod_{q=1; q \neq i}^n \frac{w_{i,l} - TID_{D_i-q}}{TID_{HSS} - TID_{D_i-q}} \mod q. \quad (11)$$

Its own e_{HSS} was also calculated:

$$e_{HSS} = g^{LC_{HSS}} \mod p. \quad (12)$$

It generates a random value $r_{D_i} \in Zq *$. The devices send TID_{D_i} , e_{D_i} , r_{HSS} , and a timestamp T_{D_i-2} to the other devices in the relay group and to HSS, which also sends TID_{HSS} , e_{HSS} , T_{HSS-2} , and r_{HSS} to the other devices in the relay group. After receiving such parameters, the entities verify the validity of the timestamp to avoid denial of service (DoS) attack. They proceed with the authentication

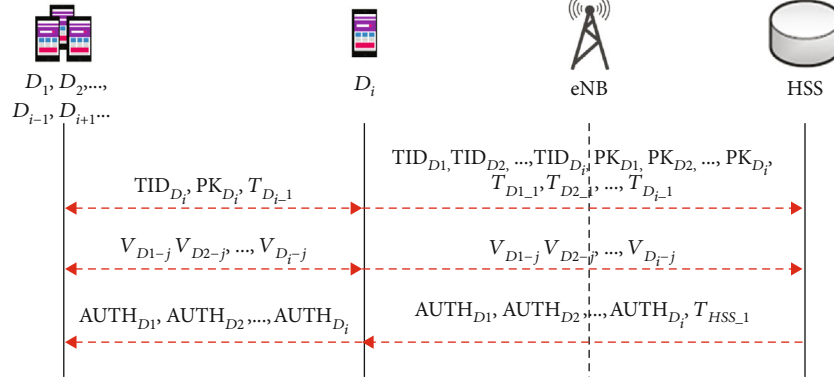


FIGURE 3: Messages exchanged in the registration phase.

only if the timestamp is valid. Otherwise, they discard the respective entity. When each entity has a complete set of e_{D_i} and e_{HSS} , a secret S' is calculated:

$$S' = \left(e_{HSS} * \prod_{i=1}^n e_{D_i} \right) \bmod p = g_i^{(LC_{HSS} + \sum LC_{D_i})} \bmod q \bmod p. \quad (13)$$

Again, an attacker must solve the DLP problem to obtain e_{D_i} , e_{HSS} , and S' , as in [15].

Next, each device checks if the $H(S')$ calculated is equal to the $H(S)$ received from HSS in the registration phase. If $H(S) = H(S')$, the devices and HSS are legit and mutually authenticated. If the verification fails, one or more intruders are in the path.

Finally, a session key is generated for each possible connection between devices D_i and HSS.

$$SK_{e_{D_i-k}} = h_3[S \| r_{D_i} \| r_u \| v_{D_i-u} \oplus v_{u-D_i}] \quad (14)$$

In this stage, if the source device has direct access to the network infrastructure, it can encrypt its health information with the session key and send it to the core network. Otherwise, it must execute phase 3.4 prior to phase 3.5.

Figure 4 shows a summary of the mutual authentication phase.

4.4. Trust Evaluation. This phase is executed whenever a patient that must send his/her health information to the doctor/physician is not in the coverage area of a 3GPP cell. Therefore, data must be relayed through other D2D devices available, until a device with a direct connection to the network infrastructure has been reached.

Due to the delicacy of the data exchanged, the trust level of each node authenticated in the mutual authentication phase must be measured before the data are sent. The trust evaluation enables the origin device to choose the path with the most reliable devices available for the relay of data. The trust system adopted follows the same idea of local trust presented by Yan et al. [12]. However, we have created our own

calculations, which differ from those of [12], due to the use of symmetric cryptography for reducing costs.

This phase is performed over an insecure channel. An architecture involving the use of relay devices, as shown in Figure 2, is employed. After the measurement of local trust, all devices considered trusted are candidates to be relay devices.

4.4.1. Local Trust Evaluation. The local trust evaluation is based on the experiences of nearby devices. Each device defines a trust threshold for deciding whether the devices are trustworthy or not.

When a device D_u wants to know if a device $D_{u'}$ is trustworthy, it compares the $LT_{D_{u'}}$ level with the desired threshold LT . If it is higher than the threshold, device $D_{u'}$ is considered trustworthy, and device D_u can relay data through it. Otherwise, the communication is refused.

Whenever a device D_u wants to obtain $LT_{D_{u'}}$ of a device $D_{u'}$, it sends $TID_{D_{u'}}$ to another device D_k , which once communicated with device $D_{u'}$, to request its local trust evaluation $LT_{D_{u'}}$. D_k generates local trust level $LT_{D_{u'}}$ of device $D_{u'}$, encrypts the result with the session key generated between D_k and $D_{u'}$, and sends it to D_u and $D_{u'}$:

$$B_{D_k-D_{u'}} = E_{SK_{D_k-D_{u'}}}(LT_{D_{u'}}). \quad (15)$$

D_u decrypts the message and obtains the local trust level of $D_{u'}$. It then checks if $LT_{D_{u'}}$ is acceptable by comparing it with the local trust threshold. If it is acceptable, D_u calculates a local trust secret key:

$$LTK_{D_{u'-u}} = h_3(S \| r_{D_i} \| r_u \| v_{D_i-u} \oplus v_{u-D_i} \| LT_{D_{u'}}). \quad (16)$$

Otherwise, it must choose another available device suitable to relay the message.

Figure 5 shows a summary of the local trust evaluation phase.

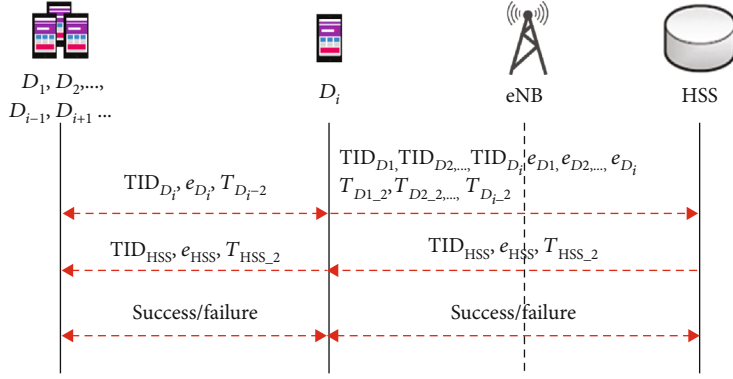


FIGURE 4: A message exchanged in the mutual authentication phase.

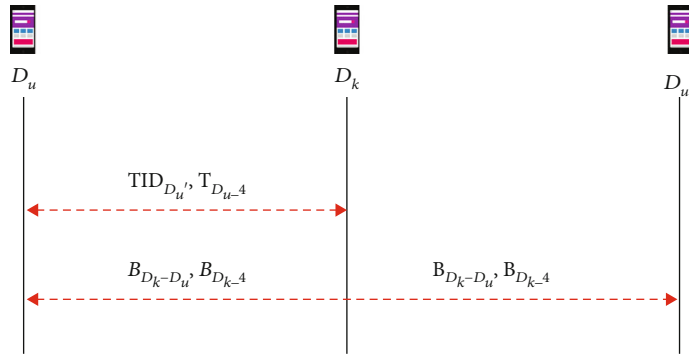


FIGURE 5: Messages exchanged when LT is used.

4.5. *Encryption/Decryption.* Finally, after the tests, the original device encrypts the data with session key $SKey_{D_i-HSS}$:

$$M = E_{SKey_{D_i-HSS}}(\text{data}). \quad (17)$$

The result (M) is encrypted with the proper key:

$$C = E_{LTK_{D_u-D_u'}}(M). \quad (18)$$

The message is sent to the most adequate device in the relay group with T_{D_i-5} and $E_{SKey_{D_u-D_u'}}(LT_{D_u'})$. Then, D_u' calculates the secret key:

$$LTK_{D_u-u'} = h_3 \left[(S || r_{D_i} || r_u || v_{D_i-u} \oplus v_{u-D_i} || LT_{D_u'}) \right], \quad (19)$$

decrypts the message, and obtains M :

$$M = D_{LTK_{D_u-u'}}(C). \quad (20)$$

D_u' encrypts M with its own trust information through Equation (17) and sends the resulting message to the most adequate device in the relay group with a timestamp T_{D_i-5} . The process is repeated until the device nearest the 3GPP infrastructure has been reached. This device sends M with a timestamp T_{D_m-5} to HSS.

HSS decrypts M using session key $SKey_{D_i-HSS}$ generated at the end of the mutual authentication phase, thus guaranteeing the legitimacy of the sender and the integrity of the data. It then forwards the patient's information to the health center server, which sends it to the doctor on a secure channel. Finally, the doctor receives the data and evaluates them.

Figure 6 shows a summary of the encryption/decryption phase, with a focus on encryption.

5. Security Analysis

This section reports on a security analysis of all D2D communication security devices and discusses the way they are approached by the proposed scheme.

5.1. *Mutual Authentication.* Devices perform mutual authentication to authenticate the other devices in the relay group. Each device calculates its Lagrange component (LC_{D_i}) and e_{D_i} , and they share e_{D_i} with the other devices in the relay group. Next, they calculate secret S' and $H(S')$ and compare the value obtained with the $H(S)$ received from HSS in the registration phase. If $H(S') = H(S)$, all devices involved are mutually authenticated. Otherwise, the operation is terminated.

After mutual authentication, the devices start the mutual authentication procedure with HSS. Each device generates MAC_{D_i} and sends it with the respective TID_{D_i} to HSS, which calculates MAC'_{D_i} and checks if $MAC'_{D_i} = MAC_{D_i}$. If the

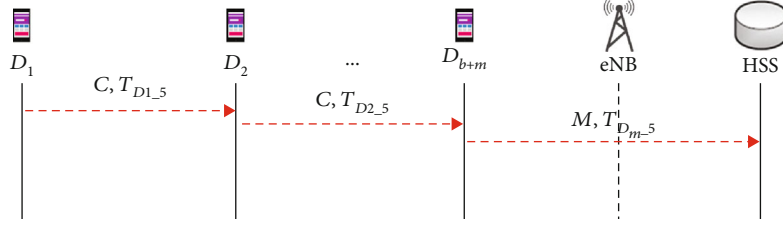


FIGURE 6: Encrypted data sent to HSS.

TABLE 2: Comparison of security objectives among protocols.

Security objectives	Zhang et al. [13]	Mustafa [18]	Zhou [16]	Proposed protocol
Mutual authentic	Yes	Yes	Yes	Yes
Trust evaluation	No	No	No	Yes
Confidentiality	No	Yes	Yes	Yes
Integrity	Yes	Yes	Yes	Yes
Privacy	No	Yes	Yes	Yes
Anonymity	Yes	No	Yes	Yes
Forward/backward secrecy	Yes	Yes	Yes	Yes
Nonrepudiation	Yes	Yes	Yes	Yes
Session key security	Yes	Yes	Yes	Yes
Resistance to replay attack	No	Yes	Yes	Yes
Resistance to insider attack	No	Yes	Yes	Yes
Resistance to DoS attack	No	Yes	No	Yes
Resistance to man-in-the-middle attack	No	Yes	Yes	Yes

TABLE 3: Cost of each unitary operation.

Notation	Devices (ms)	Network (ms)	Description
T_{hash}	0.201	0.067	Cost of a one-way hash operation
T_{mul}	1.84	0.612	Cost of a multiplication operation over an elliptical curve, represented as *
T_{add}	0.375	0.125	Cost of an addition operation over an elliptical curve
T_{mod}	0.372	0.124	Cost of a modular operation
T_{exp}	0.37	0.123	Cost of an exponential operation
T_{pair}	13.53	4.51	Cost of a bilinear pairing operation
T_{PK}	1.1	0.367	Cost of public key encryption
T_{AES}	0.483	0.161	Cost of an AES encryption/decryption operation

values are equal, HSS authenticates the devices and proceeds. Otherwise, the operation is terminated. Then, HSS generates its own Lagrange component LC_{HSS} and e_{HSS} and sends e_{HSS} to the group of relay devices. Each device recalculates its own Lagrange component $LC_{\text{new}_{D_i}}$, $e_{D_i-\text{new}}$, and a new secret S'' and compares S'' with secret S' previously calculated. If $S'' = S'$, HSS is authenticated by the devices. In the proposed scheme, an attacker finds a Lagrange component by solving the DLP problem, which has proven computationally infeasible.

5.2. Forward/Backward Secrecy of Session Key. Forward secrecy guarantees an intruder with access to an old key does not use it in the future for forging its authenticity. On the other hand, backward secrecy provides security against the use of newer keys for access to information originated in older sec-

tions. In the proposed scheme, forward and backward securities of the session key are guaranteed through the use of freshly generated random values r_{D_i} , timestamps T_{D_i} , and session keys $SK_{D_i-\text{HSS}}/SK_{D_i-D_k}$ in each authentication procedure.

5.3. Confidentiality. The scheme provides confidentiality of patients' data by generating a different session key $SK_{D_i-\text{HSS}}/SK_{D_i-D_k}$ in each session established between any device and HSS. All data exchanged over an insecure channel are encrypted with the respective session key, whose security is ensured.

5.4. Integrity. Data integrity is guaranteed by the encryption of the data sent by each patient through a securely established session key $SK_{D_i-\text{HSS}}/SK_{D_i-D_k}$ before it is sent over an insecure channel. When HSS decrypts the messages with the

TABLE 4: Comparison of computational costs.

Protocol	Devices (ms)	Server network (ms)	Total (ms)
Zhang et al. [13]	$(3n + 3m + 9)T_{\text{mul}} + 2nT_{\text{mod}} + (4m + 9)T_{\text{hash}} + 2nT_{\text{exp}}$ $+ (n + 4m + 2)T_{\text{add}} + 3T_{\text{ECC}}$ $= 7.38n + 7.83m + 22.42$	$(n + 6)T_{\text{mul}} + 2nT_{\text{mod}} + (n + 7)T_{\text{hash}} + 2nT_{\text{exp}}$ $+ (n + 2)T_{\text{add}} + 2T_{\text{ECC}} = 1.3n + 5.13$	$(4n + 3m + 15)T_{\text{mul}} + 4nT_{\text{exp}} + (n + 4m + 16)T_{\text{hash}}$ $+ 4nT_{\text{exp}} + (2n + 4m + 4)T_{\text{add}}$ $+ 5T_{\text{ECC}} = 8.68n + 7.83m + 27.53$
Zhou [16]	$(3n + 3m + 11)T_{\text{mul}} + (2n + 2)T_{\text{mod}} + (4m + 12)T_{\text{hash}}$ $+ 2nT_{\text{exp}} + (n + 4m + 2)T_{\text{add}} + T_{\text{ECC}}$ $= 7.38n + 7.83m + 25.25$	$(n + 10)T_{\text{mul}} + 2nT_{\text{mod}} + (n + 6)T_{\text{hash}} + 2nT_{\text{exp}}$ $+ (n + 5)T_{\text{add}} + T_{\text{ECC}} = 1.3n + 7.52$	$(4n + 3m + 21)T_{\text{mul}} + (4n + 2)T_{\text{mod}} + (n + 4m + 18)T_{\text{hash}}$ $+ 4nT_{\text{exp}} + (2n + 4m + 7)T_{\text{add}}$ $+ 2T_{\text{ECC}} = 8.68n + 7.83m + 32.77$
Proposed protocol	$nT_{\text{mul}} + 3nT_{\text{mod}} + (n + 2m)T_{\text{hash}} + nT_{\text{exp}}$ $+ 3nT_{\text{ECC}} + (3m + 1)T_{\text{AES}}$ $= 6.83n + 1.85m + 0.48$	$(n + 1)T_{\text{mul}} + (2n + 7)T_{\text{mod}} + 3nT_{\text{hash}} + (2n + 4)T_{\text{exp}}$ $+ nT_{\text{ECC}} + T_{\text{AES}} = 1.67n + 2.13$	$(n + 1)T_{\text{mul}} + (5n + 7)T_{\text{mod}} + (4n + 3m)T_{\text{hash}}$ $+ (3n + 4)T_{\text{exp}} + 4nT_{\text{ECC}} + (4m + 2)T_{\text{AES}}$ $= 8.5n + 1.85m + 2.61$

appropriate session key, it knows the information was generated by the genuine source and not modified on the way to the destination.

5.5. Anonymity. The anonymity of entities, devices, and HSS is safeguarded through the exchange of only temporary identities (TID_{D_i} , TID_{HSS}) over an insecure channel. Therefore, the permanent identities are not disclosed over an insecure channel. HSS knows the permanent identity of all devices; however, this information is obtained offline.

5.6. Nonrepudiation. Nonrepudiation certifies an entity cannot deny its actions. In the proposed scheme, it is guaranteed through the use of permanent (ID_{D_i}) and temporary identities (TID_{D_i} , TID_{HSS}) and private and public keys.

5.7. Session Key Security. The security of the session key is ensured by confidential information Y_{D_i} and ID_{D_i} in its generation process, as shown in Equation (13).

5.8. Resistance to Impersonation Attack. Impersonation attack is avoided by different temporary identities in each session established. A TID is never used twice and HSS can recognize whether a certain TID has already been used.

5.9. Resistance to Replay Attack. A replay attack is avoided by freshly generated parameters, such as random values r_{D_i} and timestamps T_{D_i} in the mutual authentication phase, generation of session keys, and use of different TID_{D_i} and TID_{HSS} in each session.

5.10. Resistance to Denial of Service (DoS) Attack. The use of timestamps in each message exchanged over an insecure channel avoids denial of service (DoS) attacks. Each timestamp is synchronized with its respective entity's clock, which is also synchronized with the whole system.

5.11. Resistance to Man-in-the-Middle Attack. Session keys and local trust keys do not depend only on values exchanged on an insecure channel, but on secret values securely exchanged in the registration phase encrypted with devices' public key.

Such security objectives were not accomplished by [13, 16 and [18]. First, any of the compared protocol performs the trust evaluation of relay devices. Secondly, as shown by [14], the scheme proposed in [13] is vulnerable to an insider attack, which compromises its confidentiality and might also affect patients' privacy and the protocol's resistance to replay and man-in-the-middle attacks.

The schemes proposed by [13, 16] are vulnerable to DoS attacks, since they do not use verification values as nonces or timestamps prior to the execution of more complex calculations. The scheme designed by [18] does not protect the anonymity of devices because it does not mention the use of temporary or pseudoidentity instead of their permanent identities.

Table 2 shows a comparison among our scheme and those of [13, 18] and [16].

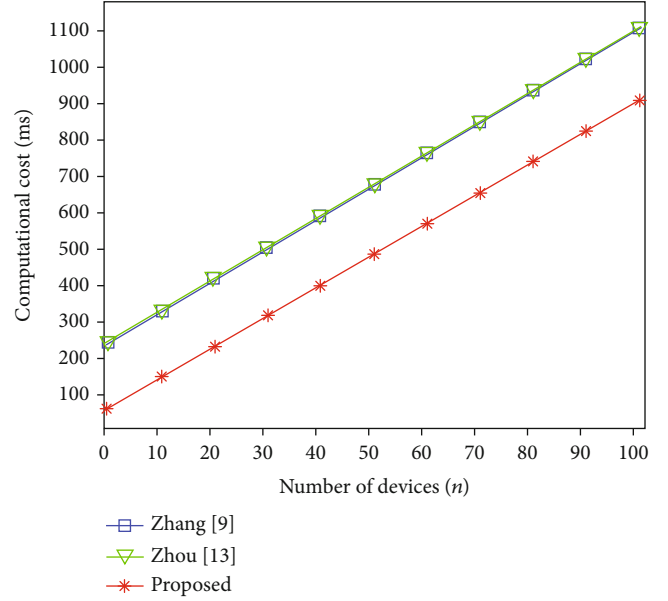


FIGURE 7: Comparison of computational costs.

6. Performance Analysis

This section reports on a performance analysis of our protocol regarding computational and energy costs in each authentication session executed, and a performance comparison among our scheme and those of [13, 16].

6.1. Computational Costs. The evaluation of computational costs is based on an estimate of the time necessary for the execution of unitary operations, considered part of the messages previously described in the different phases of the protocol. Table 3 shows such unitary operations (hash, multiplication, and addition over elliptical curve, exponentiation,...) with the corresponding computational costs (running times, measured in milliseconds) for both devices and the fixed part of the cellular wireless network.

The cost values are based on common and realistic values obtained by experimentation and adopted for performance comparisons of authentication protocols, as in Choi et al. 2014 [8] and Hsu et al. 2018 [21].

The computational platform was configured as follows:

- (i) Intel Core Duo 1.86 GHz and 2 gigabyte RAM under an Ubuntu 11.10 operating system [8]
- (ii) HTC One X smartphone with Android 4.1.1, 1.5 GHz Quad-core ARM Cortex-A9 CPU, 1 GB RAM [21]

The methodology adopted for the performance evaluation considers the cost of each unitary operation multiplied by the number of times each operation is executed, comprehending the several messages that include one or more of such unitary operations, as required for the different authentication protocols.

Table 4 shows a comparison of the computational costs (in milliseconds) among our protocol and those designed by [13, 16]. An environment with “ n ” devices

TABLE 5: Comparison of energy costs.

Protocol	Devices	Network	Total (mJ)
Zhang et al. [13]	$80.3n + 85.2m + 243.93$	$14.14n + 55.81$	$94.44n + 85.19m + 299.53$
Zhou [16]	$80.3n + 85.2m + 274.72$	$14.14n + 81.82$	$94.44n + 85.19m + 356.54$
Proposed protocol	$74.31n + 20.13m + 5.22$	$18.17n + 23.17$	$92.48n + 20.13m + 28.4$

registered in the 3GPP network and “ m ” devices involved in the relay of the messages sent from the source device and the HSS was considered. Only the devices involved in the relay of data performed the calculations of the trust evaluation phase.

The devices take $nT_{\text{mul}} + 3nT_{\text{ECC}}$ in the registration phase to calculate their partial public key and encrypt/decrypt secret values v_{D_i, D_u} . Then, they take $3nT_{\text{mod}} + nT_{\text{hash}} + nT_{\text{exp}}$ in the mutual authentication and key agreement phase to calculate their Lagrange component, secret S' , and session key $SKey_{D_i, k}$. Next, $mT_{\text{hash}} + mT_{\text{AES}}$ is required for the encryption of local trust result LT_{D_u}' and local trust secret key LTK_{D_u-u}' is calculated. Finally, the devices expend $mT_{\text{hash}} + (2m + 1)T_{\text{AES}}$ to encrypt the patients' information generating M and encrypting/decrypting M with local trust secret key LTK_{D_u-u}' . Consequently, the total computational cost for the devices is $nT_{\text{mul}} + 3nT_{\text{mod}} + (n + 2m)T_{\text{hash}} + nT_{\text{exp}} + 3nT_{\text{ECC}} + (3m + 1)T_{\text{AES}}$. According to Table 4, the computational cost for the devices is $6.827n + 1.851m + 0.483$ ms.

3GPP network takes $(n + 1)T_{\text{mul}} + 2nT_{\text{hash}}$ to calculate temporary identities and partial private keys for each device and its master public key in the initialization phase. It takes $(2n + 4)T_{\text{mod}} + (2n + 3)T_{\text{exp}} + nT_{\text{ECC}}$ to generate tokens for each device and for itself in the registration phase. Next, it requires $3T_{\text{mod}} + nT_{\text{hash}} + T_{\text{exp}}$ to calculate the Lagrange component of HSS, secret S' , and session keys $SKey_{D_i, \text{HSS}}$. Finally, it takes T_{AES} to decrypt message M and obtain the source patient's information. Therefore, the computational cost for the core network is $(n + 1)T_{\text{mul}} + (2n + 7)T_{\text{mod}} + 3nT_{\text{hash}} + (2n + 4)T_{\text{exp}} + nT_{\text{ECC}} + T_{\text{AES}}$. According to the operation costs in Table 4, the computational cost for the network is $1.674n + 2.133$ ms.

The lines in Figure 7 show satisfactory results of our protocol regarding computational costs. A situation in which 25% of devices are involved in the relay of data was considered. The protocol clearly showed better costs than [13, 16]; [16] yielded slightly different results from [13].

Regarding the similarity involving the results of [13, 16] is an improvement of [13], and, consequently, most calculations are similar.

The main difference between [13, 16] is the correction of security vulnerabilities, made in [16] with the use of more variables in the authentication procedure. In terms of operations, [16] requires the calculation of only an extra elliptic curve cryptography-based (ECC-based) scalar multiplication on G_1 when compared to [13]; such operation requires small processing time.

Our scheme has shown excellent computational performance regarding all subjects addressed. The use of Shamir's

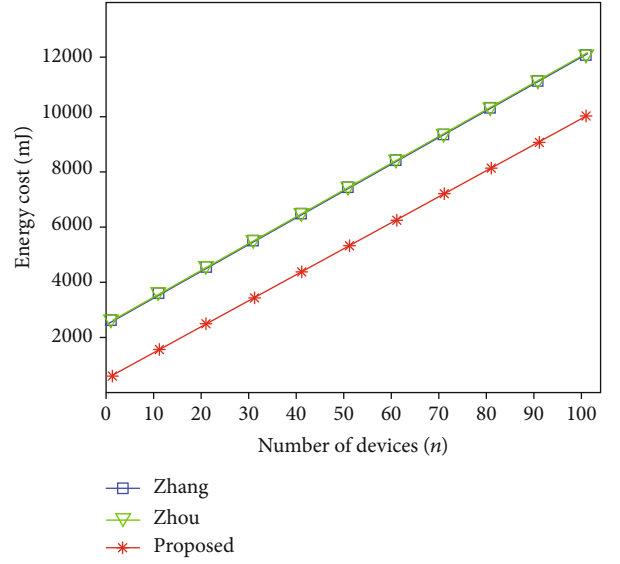


FIGURE 8: Comparison of energy costs.

secret sharing in the authentication phase reduces the computational resources consumption, since all devices and HSS are authenticated with a single calculation and comparison of secrets S' and S , respectively.

6.2. Energy Costs. This section reports on an analysis of the energy cost of our protocol and a comparison with [13, 16]. The evaluation is based on the proposals of Kumar et al. [22] and He et al. [23], which consider 10.88 Watts the maximum CPU power of devices (W). The following operation was performed for the calculation of the energy overhead: $E_{\text{Total}} = \text{CCTotal} \times W$, where CCTotal is the computational cost of each operation. Table 5 shows the results.

Figure 8 shows a comparison based on the energy costs provided in Table 5. The red line representing our protocol proves its lower energy consumption in comparison to the protocols of [13, 16]. As in the evaluation of computational costs, the energy costs of [13, 16] were similar.

The main differences between [13, 16] that lead to consequences in the energy costs are the same of the computational costs. Our scheme also shows higher energy efficiency due to reduced processing efforts and computational cost.

7. AVISPA Verification

The protocol was validated by Automated Validation of Internet Security Protocols and Applications (AVISPA) [24], which simulates messages exchanged among entities involved in an

```

role
role_Dkj(Dkj:agent,Dij:agent,HSS:agent,TIDdij:text,TIDdkj:text,TIDhss:text,PKdkj:text,Vdkdj
:text,
  AUTHdk:text,Rdk:text,Edk:text,Tdkdu:text,M:text,Ydk:text,Tdkhss:text,C:text,
  Key_set_Dkj_HSS:(symmetric_key) set,Key_set_HSS_Dkj:(symmetric_key)
set,SND,RCV:channel(dy))
played_by Dkj
def=
  local
  State:nat,PKdij:text,Vdidj:text,Fkj:text,Wkj:text,Ddkj:text,Vhssdk:text,Rdi:text,Edi:text,
  Tdkj:text,
  Rhss:text,Thss:text,Ehss:text,LTdkdu:text,Tdij:text,Data:text,Tdku:text,SecureChannel:symmet
ric_key,
  Key_4:symmetric_key,Key_3:symmetric_key,Key_2:symmetric_key,Key_1:symmetric_key
  init
  State := 0

  1. State=0 /RCV(Tdij.TIDdij.PKdij) =>
    State:=1 /\Tdkj:=new() /\SND(Tdkj.TIDdkj.PKdkj)
  2. State=1 /RCV({Vdidj}_SecureChannel) =>
    State:=2 /\ SND({Vdkdj}_SecureChannel) /\ Key_1:=new() /\
    Key_set_Dkj_HSS:=cons(Key_1,Key_set_Dkj_HSS) /\SND({Vdkdj}_Key_1)
  1. State=2 /\ in(Key_2,Key_set_HSS_Dkj) /\
    RCV({Wkj.Ddkj.Fkj.Vhssdk}_SecureChannel)_Key_2) =>
    State:=3 /\Key_set_HSS_Dkj:=delete(Key_2,Key_set_HSS_Dkj)
  2. State=3 /RCV(Tdij.TIDdij.Edi.Rdi) =>
    State:=4 /\ SND(Tdkj.TIDdkj.Edk.Rdk) /\ Key_3:=new() /\
    Key_set_Dkj_HSS:=cons(Key_3,Key_set_Dkj_HSS) /\
    SND({Tdkj.TIDdkj.Edk.Rdk}_Key_3)
  14. State=4 /\in(Key_4,Key_set_HSS_Dkj) /\
    RCV({Thss.TIDhss.Ehss}_Key_4) =>
    State:=5 /\Key_set_HSS_Dkj:=delete(Key_4,Key_set_HSS_Dkj)
  15. State=5 / \ RCV(Tdij.TIDdij) =>
    State:=6 /\LTdkdu:=new() /\Tdku:=new() /\SND(Tdku.LTdkdu) /\
    SND(Tdku.LTdkdu)
  18. State=6 /\ RCV(Tdij.{Data}_SecureChannel) =>
    State:=7 /\ secret(Data,sec_8,{})/ \ SND(Tdku.{Data}_SecureChannel) /\
    secret(C,sec_7,{})/ \ SND(Tdkhss.{C}_SecureChannel)

end role

```

FIGURE 9: Role of a device in HLPSL language for AVISPA software.

authentication scheme. AVISPA simulation is written in High-level Protocol Specification Language (HLPSL), which divides the message exchanged into roles representing each entity involved in the authentication procedure.

Figure 9 shows an example of the role of an ordinary D2D communication device. The objectives verified were ability of the protocol to perform D2D mutual authentication and key agreement and secrecy of session keys.

AVISPA has four backends to verify security. We used two of them, namely, On-the-Fly-Model-checker (OFMC) [25] and Constraint Logic-Based Attack Searcher (CL-AtSe) [26], which return “SAFE” if the protocol analyzed is considered safe, and “UNSAFE” if it has found an issue that might compromise security. According to Figures 10 and 11, the protocol was considered safe.

8. Conclusions and Future Work

Internet of Things (IoT) devices have been designed for several new applications and creation of a framework of benefits that improves services and people’s life quality, assures safety and security, and reduces expenses [27]. Some of such applications include solutions for m-health, which enable patients to share information on their health to be monitored or receive fast aid in emergencies, thus improving the quality

```

% OFMC
% Version of 2006/02/13
SUMMARY
SAFE
DETAILS
BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
/home/span/span/testsuite/results/hlpslGenFile.if
GOAL
as_specified
BACKEND
OFMC
COMMENTS
STATISTICS
parseTime: 0.00 s
searchTime: 0.72 s
visitedNodes: 63 nodes
depth: 8 plies

```

FIGURE 10: OFMC analysis.

of care [28]. D2D communication is suitable for m-health IoT applications, since it provides direct communication among devices with no intermediation of infrastructures, such as the one available by 3GPP.

The traditional authentication and key agreement standardized by 3GPP is not suitable for D2D authentication

SUMMARY
SAFE
DETAILS
BOUNDED_NUMBER_OF_SESSIONS
TYPED_MODEL
PROTOCOL
/home/span/span/testsuite/results/hlpslGenFile.if
GOAL
As Specified
BACKEND
CL-AtSe
STATISTICS
Analyzed : 77 states
Reachable : 21 states
Translation: 1.50 seconds
Computation: 0.00 seconds

FIGURE 11: CL-AtSe analysis.

and therefore cannot deal with the lack of access to the network infrastructure faced by some devices. New applications that exchange critical data (e.g., m-health applications) require novel AKA schemes to fulfill such a demand. A good alternative is the relay of data through close devices until the network infrastructure has been reached, as proposed by [12].

Our protocol has been designed to provide a new AKA scheme; it aims at fulfilling the security properties detailed by 3GPP specifications TS 23.303 [19] and TS 33.303 [11] and reducing resource consumption. Such a reduction has been achieved by the scheme adopted, as proposed by Harn [15], based on Shamir's secret sharing [17]. A trust evaluation indicated the close devices suitable for the relay of data; it was based on the scheme developed by [12], to guarantee the delivery of data from the source device to the health center

Several security properties and resistance to attacks, as addressed in Section 5.4, demonstrated the robustness of our protocol, which has proven safer than those of [13, 16]. The protocol designed by [13] showed confidentiality issues and, consequently, is not resistant to attacks (e.g., insider and man-in-the-middle). The scheme of [16] is not resistant to DoS attack, and the one developed by [18] shows anonymity problems, since it offers no protection to devices' real identities.

Our protocol has proven the safest, because it has fulfilled all security objectives required by [11, 19], as shown in Table 2, and achieved better performance, in comparison to [13, 16], which have similar costs due to their similarity. The validation made by AVISPA with the use of two of its backends also confirmed the safety of the protocol regarding message exchange of secret parameters. Therefore, no intruder can discover confidential and critical parameters and information.

Our protocol is part of a project that involves the development of software applications, which benefit from the junction of D2D communications and edge computing.

Among the application areas considered are smart cities, e-health/m-health, and smart grid networks.

Future work will include the proposal of authentication and authorization protocols based on cyber-physical systems ([28, 29]), as well as the formal validation of our protocol by tools, such as ProVerif and Tamarin. The simulation of our protocol by NS-3 or OMNET++ tools has been considered for the evaluation of energy efficiency and influence of device mobility.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

On behalf of all authors, the corresponding author states there is no conflict of interest.

Acknowledgments

This study was partially financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil (Finance Code 001) (a scholarship was awarded to Ana Paula G. Lopes). The authors acknowledge the University of Brasilia, especially the Post-Graduation Program in Electrical Engineering (PPGEE), for the research support.

References

- [1] C. Huang, K. Yan, S. Wei, and D. H. Lee, "A privacy-preserving data sharing solution for mobile healthcare," in *2017 International Conference on Progress in Informatics and Computing (PIC)*, pp. 260–265, Nanjing, China, December 2017.
- [2] M. Wang and Z. Yan, "A Survey on Security in D2D Communications," *Mobile Networks and Applications*, vol. 22, no. 2, pp. 195–208, 2017.
- [3] N. Hassan, K.-L. A. Yau, and C. Wu, "Edge Computing in 5G: A Review," *IEEE Access*, vol. 7, pp. 127276–127289, 2019.
- [4] Y. He, J. Ren, G. Yu, and Y. Cai, "D2D Communications Meet Mobile Edge Computing for Enhanced Computation Capacity in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1750–1763, 2019.
- [5] D. Wu, F. Wang, X. Cao, and J. Xu, "Joint communication and computation optimization for wireless powered mobile edge computing with D2D offloading," *Journal of Communications and Information Networks*, vol. 4, no. 4, pp. 72–86, 2019.
- [6] S. Kekki, W. Featherstone, Y. Fang et al., *ETSI White Paper No. 28 MEC in 5G networks First edition*, European Telecommunications Standards Institute, 2018.
- [7] M. Haus, M. Waqas, A. Y. Ding, Y. Li, S. Tarkoma, and J. Ott, "Security and Privacy in Device-to-Device (D2D) Communication: A Review," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1054–1079, 2017.
- [8] D. Choi, H.-K. Choi, and S.-Y. Lee, "A group-based security protocol for machine-type communications in LTE-advanced," *Wireless Networks*, vol. 21, no. 2, pp. 405–419, 2015.

- [9] J. Cao, M. Ma, and H. Li, "GBAAM: group-based access authentication for MTC in LTE networks," *Security and Communication Networks*, vol. 8, no. 17, pp. 3282–3299, 2015.
- [10] Y. Sun, J. Cao, M. Ma, H. Li, B. Niu, and F. Li, "Privacy-Preserving Device Discovery and Authentication Scheme for D2D Communication in 3GPP 5G HetNet," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, pp. 425–431, Honolulu, HI, USA, February 2019.
- [11] 3GPP TS 33.303, "Proximity-based services (ProSe); security aspects," 2018, https://www.3gpp.org/ftp/Specs/archive/33_series/33.303/. Visited on November 2019.
- [12] Z. Yan, H. Xie, P. Zhang, and B. B. Gupta, "Flexible data access control in D2D communications," *Future Generation Computer Systems*, vol. 82, pp. 738–751, 2018.
- [13] A. Zhang, L. Wang, X. Ye, and X. Lin, "Light-weight and robust security-aware D2D-assist data transmission protocol for mobile-health systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 3, pp. 662–675, 2017.
- [14] C. Zhou, "Comments on 'Light-Weight and Robust Security-Aware D2D-Assist Data Transmission Protocol for Mobile-Health Systems'," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1869–1870, 2018.
- [15] L. Harn, "Group authentication," *IEEE Transactions on Computers*, vol. 62, no. 9, pp. 1893–1898, 2013.
- [16] C. Zhou, "An improved lightweight certificateless generalized signcryption scheme for mobile-health system," *International Journal of Distributed Sensor Networks*, vol. 15, no. 1, Article ID 155014771882446, 2019.
- [17] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 2001.
- [18] U. Mustafa and N. Philip, "Group-Based Key Exchange for Medical IoT Device-to-Device Communication (D2D) Combining Secret Sharing and Physical Layer Key Exchange," in *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, London, UK, January 2019.
- [19] 3GPP TS 23.303, "Proximity-based services (ProSe); stage 2," Jun. 2018. https://www.3gpp.org/ftp/Specs/archive/23_series/23.303/. Visited on November 2019.
- [20] 3GPP TS 36.843, "Technical specification group radio access network; study on LTE device to device proximity services; radio aspects," 2014, https://www.3gpp.org/ftp/Specs/archive/36_series/36.843/, Visited on November 2019.
- [21] R. H. Hsu, J. Lee, T. Q. S. Quek, and J. C. Chen, "GRAAD: Group Anonymous and Accountable D2D Communication in Mobile Networks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 449–464, 2018.
- [22] A. Kumar and H. Om, "Handover authentication scheme for device-to-device outband communication in 5G-WLAN next generation heterogeneous networks," *Arabian Journal for Science and Engineering*, vol. 43, no. 12, pp. 7961–7977, 2018.
- [23] D. He, S. Chan, and M. Guizani, "Handover authentication for mobile networks: security and efficiency aspects," *IEEE Network*, vol. 29, no. 3, pp. 96–103, 2015.
- [24] The AVISPA Project, *European Union in the Future and Emerging Technologies (FET Open)* <http://www.avispa-project.org/> Visited on November 2019.
- [25] D. Basin, S. Mödersheim, and L. Viganò, "OFMC: a symbolic model checker for security protocols," *International Journal of Information Security*, vol. 4, no. 3, pp. 181–208, 2005.
- [26] M. Turuani, "The CL-Atse protocol analyser," in *Term Rewriting and Applications. RTA 2006. Lecture Notes in Computer Science, vol 4098*, F. Pfenning, Ed., pp. 277–286, Springer, Berlin, Heidelberg, 2006.
- [27] C. Free, G. Phillips, L. Felix, L. Galli, V. Patel, and P. Edwards, "The effectiveness of M-health technologies for improving health and health services: a systematic review protocol," *BMC Research Notes*, vol. 3, no. 1, p. 250, 2010.
- [28] A. Essa, T. Al-Shoura, A. Al Nabulsi, A. R. Al-Ali, and F. Aloul, "Cyber physical sensors system security: threats, vulnerabilities, and solutions," in *2nd International Conference on Smart Grid and Smart Cities (ICSGSC)*, Kuala Lumpur, Malaysia, August 2018.
- [29] L. Vegh, "Cyber-physical systems security through multi-factor authentication and data analytics," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, Lyon, France, February 2018.

Research Article

Smart River Monitoring Using Wireless Sensor Networks

Kofi Sarpong Adu-Manu ^{1,2}, **Ferdinand Apietu Katsriku**², **Jamal-Deen Abdulai**²,
and **Felicia Engmann**³

¹Department of Computer Science, Valley View University, Accra, Ghana

²Department of Computer Science, University of Ghana, Accra, Ghana

³School of Technology, Ghana Institute of Management and Public Administration, Ghana

Correspondence should be addressed to Kofi Sarpong Adu-Manu; kamsarpong@vvu.edu.gh

Received 27 March 2020; Revised 29 July 2020; Accepted 26 August 2020; Published 21 September 2020

Academic Editor: Hideyuki Takahashi

Copyright © 2020 Kofi Sarpong Adu-Manu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Water quality monitoring (WQM) systems seek to ensure high data precision, data accuracy, timely reporting, easy accessibility of data, and completeness. The conventional monitoring systems are inadequate when used to detect contaminants/pollutants in real time and cannot meet the stringent requirements of high precision for WQM systems. In this work, we employed the different types of wireless sensor nodes to monitor the water quality in real time. Our approach used an energy-efficient data transmission schedule and harvested energy using solar panels to prolong the node lifetime. The study took place at the Weija intake in the Greater Accra Region of Ghana. The Weija dam intake serves as a significant water source to the Weija treatment plant which supplies treated water to the people of Greater Accra and parts of Central regions of Ghana. Smart water sensors and smart water ion sensor devices from Libelium were deployed at the intake to measure physical and chemical parameters. The sensed data obtained at the central repository revealed a pH value of 7. Conductivity levels rose from 196 S/cm to 225 S/cm. Calcium levels rose to about 3.5 mg/L and dropped to about 0.16 mg/L. The temperature of the river was mainly around 35°C to 36°C. We observed fluoride levels between 1.24 mg/L and 1.9 mg/L. The oxygen content rose from the negative DO to reach 8 mg/L. These results showed a significant effect on plant and aquatic life.

1. Introduction

Monitoring and communicating water quality data in real-time is critical for promoting sustainable development. Rivers are vital resources that support the life of both humans and animals. Hence, their importance cannot be overlooked especially in developing countries where there are several reports of water-borne diseases [1, 2]. Water quality monitoring networks are required to monitor rivers efficiently. The goal is to alert stakeholders and the citizenry about changes in river water quality. Traditionally in Ghana, trained field officers are sent to the field to collect water samples for analysis in the lab. This approach is challenging, time-consuming and expensive, and may not yield the required temporal granularity that is needed [3, 4]. New devices capable of in situ measurements have been proposed to overcome some of the challenges associated with traditional methods of gathering samples [4, 5].

Examples of such devices included pH meter HANNA-pH211, HACH-2100 P turbidity meter, and HACH-Dr 5000 spectrophotometer, which are used mostly to measure some of the physiochemical parameters of the water. These devices could not ensure reliable and timely data collection during the water quality monitoring process since they are not deployed permanently in the water but have to be taken to site when required and in most cases, this may be well after the event [6, 7].

A modern approach is to make the devices autonomous by equipping them with wireless transceivers and form them into a wireless sensor network (WSN). WSNs thus comprise of autonomous, self-configuring and battery-powered sensor nodes and base station nodes (or sinks) capable of measuring some physical phenomenon and transmitting the information acquired to a central office. The sensor devices, apart from sense, are capable of performing computations and

communicating via different network communication protocols (such as ZigBee, WiFi, WiFiDirect, GSM, and WiMax) [3]. The sensor nodes are usually small, low-cost devices suitable for a variety of applications ranging from environmental to security applications which do not require any existing infrastructure to operate [8]. These sensors may be fixed at a permanent location or allowed to move within the environment to measure the parameters of interest.

The spatiotemporal characteristics of water may pose unique challenges for WSN deployment. In freshwater bodies, the sensor nodes deployed at different sampling locations from where the data is collected and preanalysed before transferring to a remote central office. Communication is achieved between the sensor node and the remote central office using a low-power radio transmitter in order to maximise the lifetime of the sensor nodes. The use of wireless sensor networks comes with advantages such as autonomy, reliability, robustness, and flexibility, speed, accuracy, and cost-effectiveness [9]. These advantages make WSN a preferred technology to monitor the presence of pollutants in river bodies and gather a substantial amount of information for policy implementation and managerial planning. In most deployments, the data gathered by the sensor nodes is sent through a gateway to the cloud or a database using an existing communication network. The sensed data from the sensors are made available on a web portal in real time, enabling stakeholders to intervene in the water quality management processes effectively.

In this paper, we demonstrate a practical solution to the water quality monitoring challenges in Ghana. Libelium wireless sensors were deployed at the intake to the Weija dam in Ghana to obtain continuous real-time data which is made available to the public via a web portal. The accuracy and precision of the parameters recorded by the smart sensor nodes are compared to the measurements obtained by the field personnel in the laboratory. The choice of using Libelium sensors in this project was influenced by (1) robust sensor construction (sensor modules are housed in a waterproof case); (2) high accuracy in sensor readings; (3) simple setup process and ease of use; (4) cost of sensors and maintenance warranty policy; (5) low energy consumption; (6) reliability; (7) online and offline support through webinars, product training, discussion forum, and technical inquiries through email; and (8) software compatibility with other systems such as Arduino sensor board. These meet the criteria set for selecting sensor devices for water quality monitoring [10]. The idea is to benchmark the performance of these sensors to enable us to compare to fibre optic sensors which will be designed and deployed in the future project.

The rest of the paper is organised as follows. Section 2 of the paper provides detailed information on real-time deployments on smart water quality monitoring in some parts of the world. Section 3 provides information on the materials and methods used and provides background information on water quality parameters, system deployment strategy, data storage, monitoring, and transmission. Section 4 presents WSN design experimentation, results, and discussions. Section 5 provides future directions. Finally, Section 6 concludes the paper.

2. Related Works

Wireless sensor networks have been widely adopted to monitor different phenomena in the environment [11–16]. For example, WSNs are used to monitor air quality [17], water quality [3], tracking endangered species [18], and monitoring animal habitat [19]. The adoption of WSNs allows the elimination of problems associated with conventional monitoring approaches. An application area that has received attention in recent years has been water quality monitoring. Freshwater sources have been subjected to various kinds of pollution globally. Many projects worldwide have been directed at monitoring and improving freshwater sources in many regions of the world including, Australia [20], Ireland [21], China [22], Fiji Islands [23], Portugal [24], and Kosovo [25]. In most of these projects, physical parameters were measured, and data collected over a communication infrastructure to a central repository.

In Australia [20], the authors deployed wireless sensor nodes to monitor coastal underground water resources in the Burdekin area, Queensland, Australia. The sensors were deployed in the field to collect real-time water quality data and the amount of water pumped from the area from April 4, 2007, to April 18, 2007. The distance between two nodes on the average was around 850 meters. The use of long-range communication caused end-to-end delays. Hence, the nodes perform retransmission of data packets for about six tries. To overcome this, the authors used a surge (that is, a multihop routing protocol), a reliable routing protocol to improve the connectivity between nodes. In River Lee in Ireland [21], the authors provided a water monitoring system called DEPLOY. The DEPLOY project connected five (5) different zones in the city of Cork to monitor the water quality using a multisensor system in 2009. GSM was used for data transmission from the wireless sensor network backbone to the central repository. In the Fiji Islands [23], the authors designed a smart WQM system based on Internet of Things (IoT) and remote sensing (RS) techniques to improve the traditional methods for measuring water quality parameters in the Fiji Islands. The sensor nodes transmit the sensed parameters via a GSM communication module. Sensed data are read and transferred continuously for an hour, and the system is made to sleep for 15 minutes for energy conservation. The authors further extended the lifespan of the batteries used by the sensor devices by setting idle modules to off mode. In Portugal [24], a WSN project that uses a Libelium Smart Water kit for remote water quality monitoring was implemented. The nodes are designed to use 802.15.4 radios to communicate to the Meshlium (i.e., this device serves as the gateway) via 3G or GPRS communication module. The sensor devices are equipped with 6600 mAh battery and a 2 W solar panel which improved the lifetime of the sensor devices, the used of the 3G communication module (Meshlium). In the River Sitnica [25], an intelligent water monitoring system is deployed to monitor water quality parameters in Kosovo to provide datasets to policymakers, water experts, and citizens. Measurements were taken every 7 or 10 minutes, or on-demand through a gateway node to a central monitoring node through a ZigBee protocol and GPRS.

Although there are attempts to implement and deploy sensor devices to monitor and measure water quality continuously in some parts of the world, factors such as cost, energy efficiency, hardware/software issues, and communication issues challenge the implementation of such projects in developing countries [26]. This implementation to the best of our knowledge serves as one of the first real-time deployments for monitoring rivers. A critical consideration for monitoring river bodies in real time is to obtain a large amount of data for trend analysis and scientific research studies. This work, therefore, is one that showcases an application of wireless sensor networks for real time and continuous monitoring of river bodies in Africa and makes data available for scientific discourse and decision-making.

Also, we presented a new data format different from the Libelium binary data format. To minimise the energy consumed by the nodes, we reprogrammed the devices to control the data transmission and harvested solar energy to power the sensor nodes throughout the project.

3. Methods and Materials

Polluted freshwater sources threaten the existence of aquatic life (both plants and fish), humans, and animals relying on the water bodies. Several factors affect the level of chemical concentrations of physiochemical parameters in freshwater. In this study, water quality sensor probes capable of measuring pH, calcium ion (Ca²⁺), conductivity, dissolved oxygen, fluoride ion (F⁻), nitrate ion (NO₃⁻), oxidation-reduction potential, and temperature are used for detecting contaminants in a field experiment. The implementation details of the proposed approach for measuring water quality parameters are provided in this section. Detailed description of the architecture of the river water quality monitoring system, the system deployment strategy, data storage, monitoring, and transmission system are provided.

3.1. The Architecture of the River Water Quality Monitoring System (Network Architecture). The network architecture in Figure 1 shows the deployment area, the communication infrastructure, and the monitoring station where the data collected are analysed.

The Libelium Waspote (that is, the smart water kits and smart water ion kits) sensor nodes were used for the deployment. The network was designed using the star topology, as illustrated in Figure 2. Each sensor node communicates directly with the base station via the gateway node to reduce latency. The sensor hardware architecture is composed of the sensing subsystem, the processing/computing subsystem, power subsystem, and communication subsystem, as shown in Figure 3. The sensing subsystem depends on the sensor probes to sense parameters from the region of interest.

The processing subsystem is composed of the data acquisition of raw data, data processing, and data analysis. The power subsystem supplies power and monitors the power consumption of the sensor node. The primary power source of the Waspote Smart Water Sensor is a rechargeable battery. A solar panel is also attached to the power subsystem to harvest solar energy from the environment to recharge

the node's battery after depleting over time. The communication subsystem used is varied. The common types used by the smart water device include WIFI. The Libelium Waspote sensor uses cellular communication modules such as 3G/4G, General Packet Radio Service (GPRS), long-range 802.15.4/ZigBee (868/900 MHz), and Wideband Code Division Multiple Access (WCDMA) connectivity to transmit data/information to the cloud. The 4G communication model features a GPS that enables researches to perform real-time monitoring and allows the use of a SIM card for GSM network communication.

3.2. System Deployment Strategy. The section presents the system deployment strategy and focuses on the sensor probes, the calibration process, and the cloud-based web portal design used for reporting and analysing the data obtained from the deployment environment.

3.2.1. Sensor Probes. The sensor probes measure the levels of concentration of contaminants and offer several advantages. These include the following: (1) real-time information on concentrations will provide an adequate solution to the ongoing questions to the management of freshwater sources in Ghana especially at a time that the various freshwater sources are being contaminated with mining activities; (2) reporting of higher or lower concentrations will inform decision-makers about treatment levels to obtain good-quality drinking water; (3) continuous and accurate information will be provided on hourly, daily, monthly, and yearly bases about the concentration loads of the freshwater sources; (4) trend analysis; (5) development of statistical models to support realistic estimation of concentrations over some time using discrete time-series information; (6) provision of new insights for scientific studies and analysis about the ecosystem health; and (7) early warning signal [28]. To better understand the choice of the physiochemical parameters used in this experiment, we provide brief descriptions of each of the parameters in Table 1.

3.2.2. Sensor Calibration. In WSN for WQM-based applications, data obtained from the sensor nodes affect data analysis and reporting. Therefore, the sensor nodes must be handled with much care from calibration to the deployment environment. Sensor calibration is performed to reduce errors, maintain consistency and accuracy in measurements, and to improve sensor performance [29]. Manufacturers calibrate industrial sensors for environmental applications, but for real-time monitoring applications, it is recommended that sensors are recalibrated before deployment to remove the errors in data readings. Sensors become faulty or erroneous due to the following factors: (1) subjection to heat; (2) cold; (3) humidity, and (4) shock during assembling, packaging, shipment, and storage. Wireless sensor devices also are likely to lose their sensitivity over time and hence, require regular recalibration to maintain their accuracy [30]. A well-calibrated sensor will, given an input, produce the same output independent of how many times the measurement is taken (i.e., precision). The sensor node's precision is affected by noise and hysteresis.

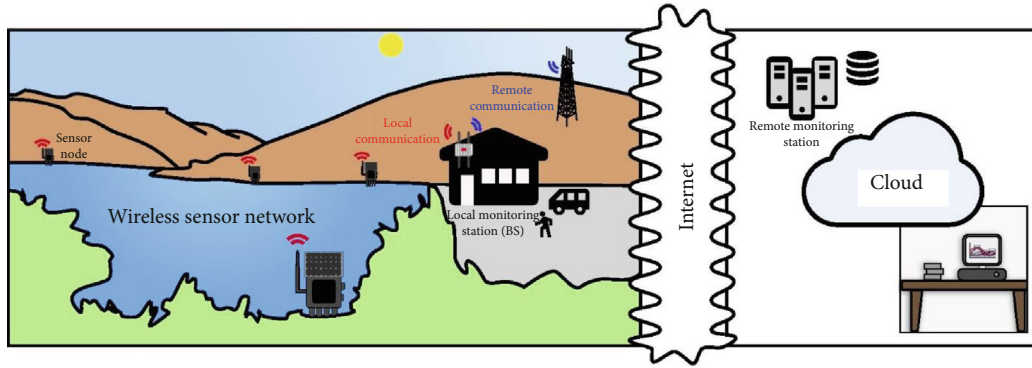


FIGURE 1: Network architecture [27].

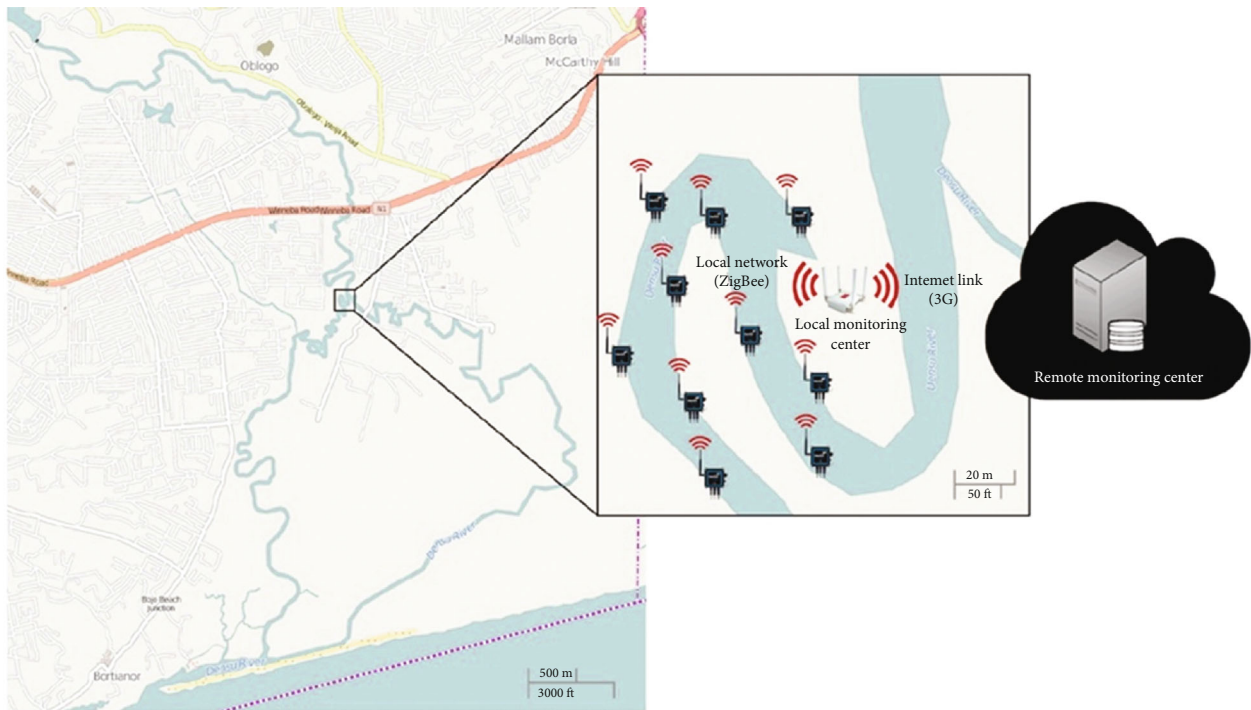


FIGURE 2: Network topology.

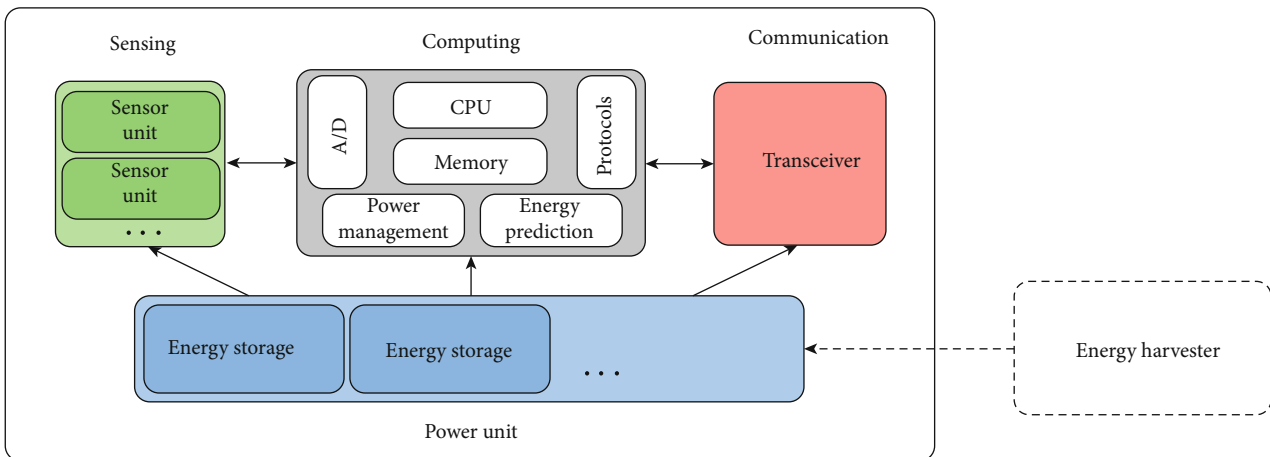


FIGURE 3: Sensor Hardware Architecture [27].

TABLE 1: Characteristics of the sensor nodes used in the project.

Sensor node/characteristics	Range/concentration	Sensitivity	Accuracy	Cable length	The temperature level of operation	pH levels
Temperature (Pt-1000)	0 to 100°C	±0.1	±0.1°C at 0°C	150 cm	-	-
pH probe	0–14 pH	±0.1	-	~500 cm	0~80°C	-
Conductivity probe	$1 \pm 0.2S/cm^{-1}$	±1	-	~500 cm	-	-
Calcium probe	0 to 100 mg/L	24 ± 5	-	-	5 to 50°C	5–8 pH
Nitrate probe	-	-54 ± 5	-	-	5 to 50°C	2–11 pH
Fluoride probe	1-25 mg/L	-54 ± 5	-	-	5 to 50°C	4–8 pH
ORP probe	0~±1999 mV	-	-	500 cm	-	-
Dissolved oxygen probe	4–7 mg/L (dry)/12-13 mg/L (rainy)	-	±2%	500 cm	50°C	-

The sensor probes are calibrated with a standard reference solution for stable sediment concentrations. Each of the probes was immersed in distilled or deionised water to clean any impurities before placing the probe in the reference solution to obtain the precise value of the cell constant. The pH sensor was calibrated to give accurate and precise pH readings. The buffer solutions were pH 4.0, 7.0, and 10.0. The pH calibration was performed at 25°C. The accuracy and precision of the probe were determined after some time to obtain the following standard pH values of 1.998, 2.099, and 2.244, respectively. A similar calibration process was undertaken to immerse all the other probes into their respective buffer solutions for some time to obtain new standard values as follows: conductivity (124 and 176), calcium (2.049, 2.430, and 3.981), and DO and ORP (2.796 and 0.00).

The sensor nodes deployed in the freshwater are affected by environmental conditions and interferences during communication which leads to noise. Interference or noise from signals affects accuracy and precision in measurements. It may be seen that hysteresis can lead to data measurement delays which may significantly affect data accuracy and precision. Hysteresis in temperature sensors, for example, may arise in an event where moisture penetrates inside the sensor node. Hysteresis may also arise if an amount of strain is applied to the sensor in freshwater monitoring [31]. For real-time water quality monitoring, wireless sensors perform continuous sensing and transfer the measured values to a storage location for analysis and reporting. Sensors with low signal-to-noise-ratio (SNR) do not support continuous water quality monitoring because such sensors have problems taking repeatable measurements. Also, when the sensor values estimated lags the changes in the river, it affects precision.

3.2.3. Cloud-Based Web Portal Design. The relevance of wireless sensor networks for water quality monitoring lies in their ability to provide up to the minute information to stakeholders. One effective way in which this can be achieved is using web portals and SMS alerts. We designed an intelligent web portal to collect, model, store, retrieve, manipulate, analyse, visualise, and share data and information obtained from the sensor devices in real time. The cloud-based web portal shown in Figure 4 is used to track the positions of the wireless sensor nodes. In smart river monitoring projects, massive amounts of data are generated from the sensing nodes.

The general architecture for the deployment is shown in Figure 5. The sensor nodes transmit data through a GSM/GPRS to the cloud. The Wasp mote data frame is designed with a specified structure. The data frame is binary or ASCII, and it also defines frames based on the activity the frame performs (i.e., event or alarm). In this paper, the data frame to be transmitted is shown in Table 2. In the frame structure, each sensor has a unique identification number, the device ID, which is 8 bytes in length. The name of the parameter is stored in the field labelled parameter name which is 16 bytes in size. The sensor type field is a 3-byte field that describes the two kinds of sensor devices used: smart sensor unit and smart water ion. A 4-byte field is assigned to the measured value from the sensor node. The data is measured in real time; hence, we keep track of the date and time the measured value was received. The size of this field is between 3 and 10 bytes. The location information (that is, the sensor device has a GPS sensor on board to measure the latitude and longitude information) is stored in the location field, which is 4 bytes.

The data acquisition system has been designed to transmit water quality data to a monitoring centre via a 4G communication infrastructure. The data values received from the sensor devices are processed and analysed directly online. The connectivity status contains the codes for the 4G connectivity attempts. The system checks for the host, the port, and resource (i.e., the location of the sensor node, the probe ID, and the reading) for availability to establish a connection; it verifies the name of the access point (that is, the APN used is Vodafone Internet), the login credentials (user name and password), the host (that is, the server IP to which the HTTP request is sent with the probe readings), the port number, the connectivity status, and the HTTP status. When these verifications are completed, and a connection is available, readings are sent to the cloud; otherwise, the 4G communication module is put to deep sleep to save energy. When a sensor node is ready to send readings, and it loses the connection, the process is repeated until a 4G connectivity is established for data to be transferred successfully.

3.3. Data Storage, Monitoring, and Transmission (Communication)

3.3.1. Data Storage. Smart water sensors and ion sensors have been deployed at the intake to monitor physical and chemical

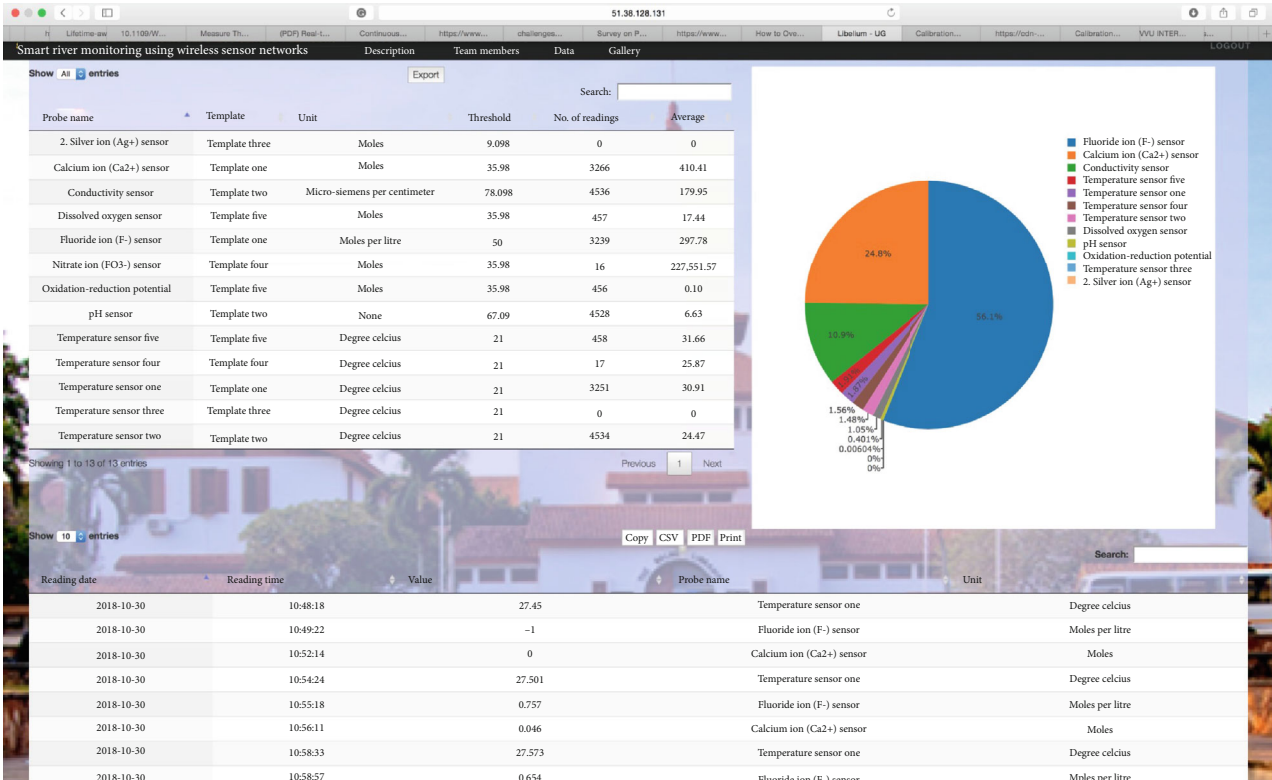


FIGURE 4: Web portal showing data received from the deployment site.

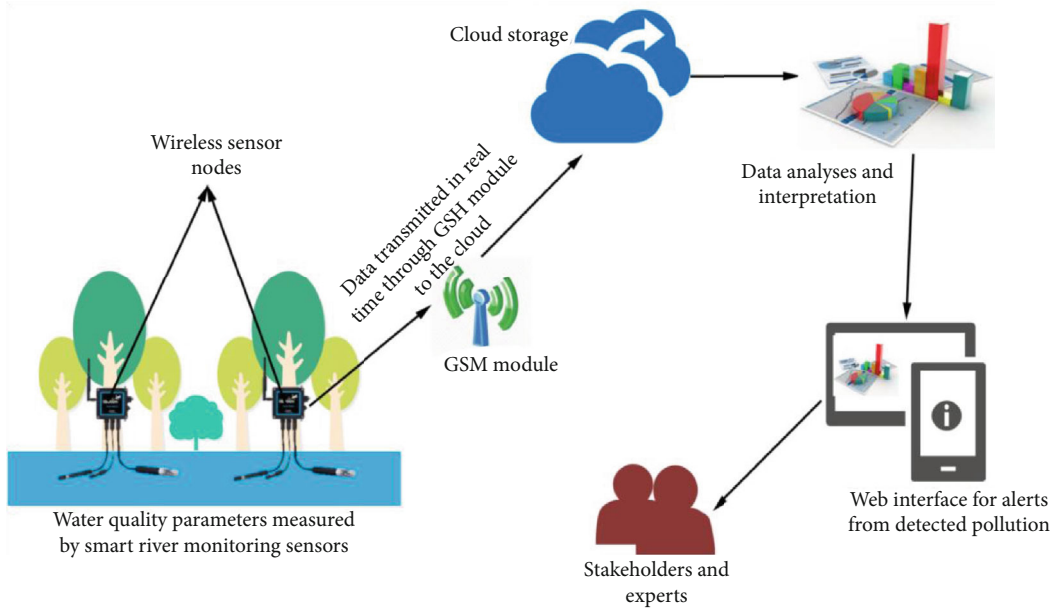


FIGURE 5: The monitoring process.

TABLE 2: Data format.

Header					Payload		
Device ID	Parameter name	Location	Sensor type	#	Measured value	Packet Rx date/time	Threshold value

parameter composition in the river body. The summary of data collected from September 30, 2018, to February 13, 2019, is presented in Table 3. The data obtained was stored in the cloud and retrieved for real-time analysis (see Figures 4 and 5). The pH, electrical conductivity, calcium, fluoride probes, and temperature of the sensor devices were programmed to transmit data packets every 30 minutes and the dissolved oxygen and the oxidation-reduction potential transmitted data packets every hour. The time interval was set to overcome communication delays and energy consumption issues. The data presented here is for a period of five (5) months from September. The data received from the probes are observed in real time from a web portal at a monitoring centre. There were some anomalies observed in the transmission of the data (see Table 4).

The data obtained from the Weija intake were observed to have an average delay of 3.22 minutes. This delay may be attributed to the distance between the sensor nodes and the base station, which is approximately 1500 meters. Other factors that may contribute to the delay include the send time, access time, and transmission time. Based on the results obtained, the average delay may be attributed to the send time, which is nondeterministic and the propagation time, which is dependent on distance.

In some sections of the data collected, due to the MAC protocol implementation, delays were observed. In Table 4, the data obtained from the conductivity sensor on some of the days in November 2018 and the month of January 2019 are recorded indicating the delay time. In some instances, over some seconds, the same data were sent either two or three times since the transmitter did not receive ACK from the base station, it forwarded the data value again after some seconds. For example, from Tables 4, 180.853 was sent after 18 seconds and 9 seconds, respectively, due to the propagation delay. In Table 5, the data obtained from the pH sensor in November 2018 is recorded. We realised that there were anomalies in the reading time, and there were indications of differences in time of receiving the data transmitted from the sensor nodes.

3.3.2. Biofouling. In water quality monitoring applications, the sensors are deployed in the freshwater source for several months and sometimes years due to the application requirements. The length of time the sensors stay immersed in the freshwater sources causes biofouling of the sensors, which affects the accuracy and data consistency, analysis, and reporting. Biofouling is the accumulation of unwanted plants, algae, microorganisms, and animals exposed on the surface, which forms around the sensor node [32]. After some months of deployment of the various sensors, there was data degrading owing to fouling of the wireless sensors (Figures 6 and 7).

The sensor nodes operating lifetime are decreased due to biofouling. Biofouling of the sensor devices generally introduces errors into the measured data values. Conventional methods for removing biofouling as such as wiper mechanisms, copper corrosion mechanisms, and chlorine evolution mechanisms are not suitable for removing fouling of sensors in this environment [3]. These existing techniques are costly

TABLE 3: Summary of data collection from September 30, 2018, to February 13, 2019.

Sensor probe	Number of readings	Average reading	Standard deviation
Calcium ion (Ca ²⁺) sensor	3266	3.13	0.455
Fluoride ion (F ⁻) sensor	3239	0.12	0.0093
Conductivity sensor	4302	186.11	70.106
pH sensor	4294	9.86	5.19
Oxidation-reduction potential	431	0.098	0.035
Dissolved oxygen sensor	432	17.36	30.439
Temperature sensor one	3251	30.91	1.258
Temperature sensor two	17	25.87	10.124
Temperature sensor five	433	31.65	1.549

TABLE 4: Readings with anomalies from conductivity sensor probe.

Reading date	Reading time	Value	Delay
2018-11-27	14:19:51	208.801	N/A
2018-11-27	14:19:59	208.801	00:00:08
2018-11-27	14:52:31	203.216	00:32:32
2018-11-14	15:48:25	188.359	N/A
2018-11-14	16:21:06	180.853	00:32:41
2018-11-14	16:21:24	180.853	00:00:18
2018-11-14	16:21:33	180.853	00:00:09
2018-11-14	16:54:09	177.836	0:32:36
2019-01-01	00:41:17	222.754	N/A
2019-01-01	00:41:26	222.754	00:00:09
2019-01-01	01:13:52	222.553	00:32:26
2019-01-01	01:46:06	222.495	00:32:14

TABLE 5: Readings with anomalies from pH sensor probe.

Reading date	Reading time	Value
2018-11-01	17:53:50	9.629
2018-11-01	18:26:13	9.808
2018-11-01	18:58:32	9.897
2018-11-01	19:30:55	9.905
2018-11-01	20:03:18	9.741
2018-11-01	20:35:44	10.412
2018-11-01	21:08:12	9.812
2018-11-01	21:40:40	10.126
2018-11-01	22:13:03	10.263
2018-11-01	22:45:20	10.175
2018-11-01	23:17:43	10.068
2018-11-01	23:50:02	10.055

and do not efficiently work for the type of sensors adopted for the monitoring process. From the initial deployment date, trends were observed in the river for variations. We observed the high levels of algae bloom and freshwater weeds which

is successful, or else the node considers the transmission as unsuccessful hence more data frames are sent.

The successful transmission of a data frame or throughput of Aloha is $T = G * e^{-2G}$, where T is the throughput and G is the mean of the Poisson distribution over transmission attempt. The maximum throughput of Pure Aloha is approximately 0.18 (18%) when G is 1/2 equal to the total transmitted frames. To mitigate the challenges of Pure Aloha, the Slotted Aloha was introduced. In Slotted Aloha, the time is divided into discrete time slots which correspond to the length of a data frame. Nodes are permitted to transmit data in the next time slot.

The throughput $T = G * e^{-G}$ is maximum when $G = 1$, which corresponds to 37% of the total transmitted packets with 26% collision. In the Libelium Wasp mote used in the application developed, the Slotted Aloha MAC protocol is implemented with a time slot of 30 minutes for each node placed in the Weija river. The nodes implemented ZigBee protocol, and therefore each node in the network knows the location of its neighbours, but communication is not permitted between them. All communication from the deployed sensor node is made with a central base station situated about 1500 meters from the nodes, and then forwarded to the cloud where remote processing and monitoring is done. In the duty cycle implemented, we consider the connected stationary network of 5 nodes with a single sink. In the initialisation stage, all nodes are assumed to be awake at deployment time. In the next working cycle, all nodes operate with a duty cycle of 0.5, which periodically turns on the radio to transmit or receive packets based on a wake-up schedule. The network assumes a star topology where all nodes transmit to the single sink.

A specific sleep/wake-up schedule is not assumed as nodes transmit to the sink in their scheduled time slots in a one-hop data communication. A sleeping node is switched to the active state when it has data to send in its time slot. The time slot has an impact on the time and energy efficiency of the duty cycle. Shorter time slot means the frequent switches, and that affects the overhead in the operations such as for opening or closing the radio connections. More extended time slot also means there will be longer times to connect with neighbouring nodes, and connection duration also lasts longer. The delay in packet delivery may be attributed to the following: send time, access time, transmission time, propagation time, reception time, and receive time of the message. Many data transmissions are likely to increase the amount of power consumed by each wireless sensor node; therefore, we reduced the number of transmissions to achieve minimum energy consumption by the nodes during the transmission (T_x) and reception (R_x) of data packets. The 4G module selected also supported the sending of SMS and email messages in the form of alerts to stakeholders and allowed users to perform HTTP and FTP requests. The 4G module can perform multisocket connections to TCP/IP and UDP/IP clients which serves as an advantage in the deployment country.

3.3.6. Energy. In WSN, a node's communication system consumes more energy compared to sensing and local data processing [26]. Replacing the sensor node batteries in the deployment in the river is a difficult task, expensive, and

inconvenient. Depleted batteries affect the project's goal of achieving continuous monitoring for months and even years. The internal battery in most cases is unable to meet the energy budget; hence, in WSN for environmental applications such as WQM, the energy budget may be improved by designing algorithms and protocols that are capable of regulating and reducing the amount of energy consumed and the frequency of data transmission when the measured parameter is transmitted to the cloud. Also, the energy budget may be compensated by harvesting energy from ambient sources to extend the lifetime of the battery during its operation [33].

In the application domain, the possible ways to harvest energy to recharge the sensor nodes include harvesting energy from the flowing river (that is, kinetic energy), radio frequency (RF) sources (that is, from sun flares, lightning, stars in space, and electromagnetic waves), and harvesting energy from the sun using either internal or external solar panels. Apart from solar energy, all the other forms of energy harvesting are challenging to implement. It is crucial to consider the energy costs per the different modules of the sensor nodes and the different algorithms and protocols that run on the nodes. Some of the sources of energy consumption include sensor node sleep/wake-up scheduling, data sensing, signal-to-electrical conversion, and signal conditioning. Duty cycle, which is the scheduling of the sleep/wake-up modes of the sensor node, also contributes to energy consumption. In the sleep mode, all operations of the sensor node are shut down and resume when the node is awake [34].

In the event of collisions or subsequent failed attempt to transmit, which results in no acknowledgement received, the traffic control signals transmitted increase thereby increasing the energy consumed. The amount of energy harvested in solar-powered wireless sensor networks is proportional to the size of the solar panels and their conversion power. The amount of energy depleted during communication could either be energy loss during Transmission (T_x) or energy loss for Reception (R_x). Energy can be harvested from river flow, but the flow should come with higher pressure (that is, turbulent) to generate power to charge the sensor nodes. Kinetic energy harvesting in rivers also requires that the sensor nodes be designed to use small turbines that move to generate electricity to charge the batteries directly.

Energy harvesting based on temperature gradient becomes a challenging technique to adopt because of variations in weather conditions in the subregion. The current smart water sensor nodes designed by Libelium allow users to harvest energy to augment the power stored in the batteries. Hence, we harvest solar energy to compliment to prolong the lifetime of the wireless sensor nodes and the lifetime of the overall sensor network (Figure 8). The highest power density a photovoltaic cell can provide on a bright sunny day is 15 mW/cm^2 [27]. The hardware architecture followed the harvest-store-use energy harvesting architecture presented in [27]. To harvest energy, we mounted the solar panel on the protective case housing the sensor device, as shown in Figure 8. The Libelium wireless sensor node's energy consumption is compared to other available commercial sensors and as indicated in Table 6. Libelium sensors consume 30 mA of energy to transmit and receive data packet



FIGURE 8: Externally mounted solar panel for energy harvesting in freshwater.

TABLE 6: Commercial sensor node energy consumption characteristics [35].

	Libelium	IMote2	SunSpot	MicaZ
Radio standard	4G (GSM)/802.15.4/ZigBee	802.15.4	802.15.4	802.15.4/ZigBee
Microcontroller	Atmel ATMEGA 1281	Marvell PXA271	ARM 920 T	ATMEGA 128
	30 microA (sleep)			
Sleep	33 microA (deep sleep)	390 microA	33 microA	15 microA
	7 microA (hibernate)			
Processing	15 mA	31–53 mA	104 mA	8 mA
Transmission	30 mA	44 mA	40 mA	19.7 mA
Reception	30 mA	44 mA	40 mA	17.4 mA
Idle	-	-	24 mA	-
Supply	6–12 V	3.2 V	4.5–5.5 V	2.8 mW

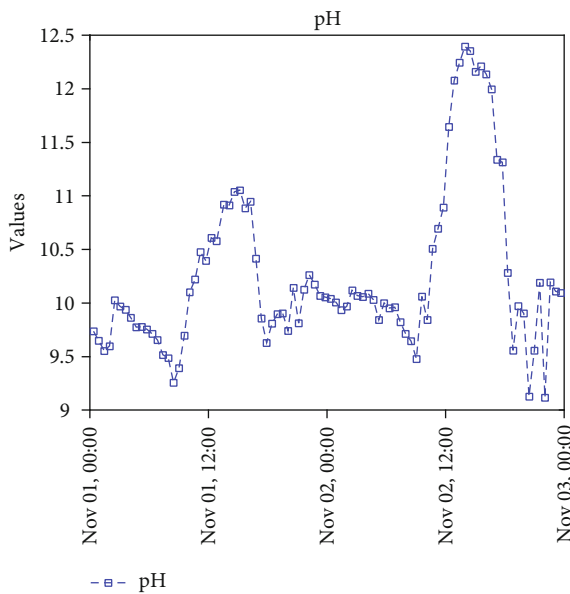


FIGURE 9: pH sensor readings from November 1 to 3, 2018.

compared to IMote2 and SunSpot that require 44 mA and 40 mA, respectively, for packet transmission and reception.

The sources of energy depletion in duty cycling could be attributed to the frequent switches between sleep and wake-up states. The frequent switches could introduce more collisions and the traffic control signals. This happens because when no acknowledgement is received or there are failed attempts to transmit, more packets are resent. The energy consumed when l -bits of data is sent over a distance d is obtained from

$$E_{Tx}(l, d) = E_{Tx-elec}(l) + e_{Tx-amp}(l)d^n, \quad (1)$$

$$E_{Tx}(l, d) = E_{Tx-elec}(l) + e_{Tx-amp} * l * d^2.$$

The energy consumed for receiving l -bits of data E_{Rx} is illustrated in Equation 2.

$$E_{Rx}(l) = E_{Rx-elec}(l), \quad (2)$$

where e_{amp} and E_{elec} are the energy dissipated per bit in the transmitter circuitry and the energy dissipated in transmitting l -bits over a distance d , respectively.

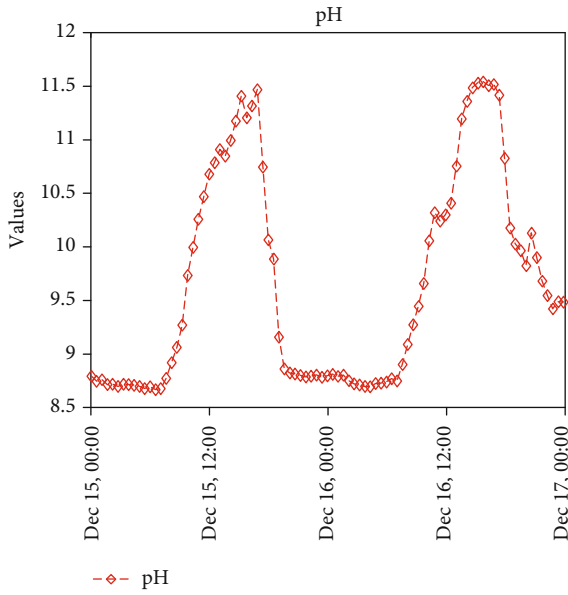


FIGURE 10: pH sensor readings from December 15 to 17, 2018.

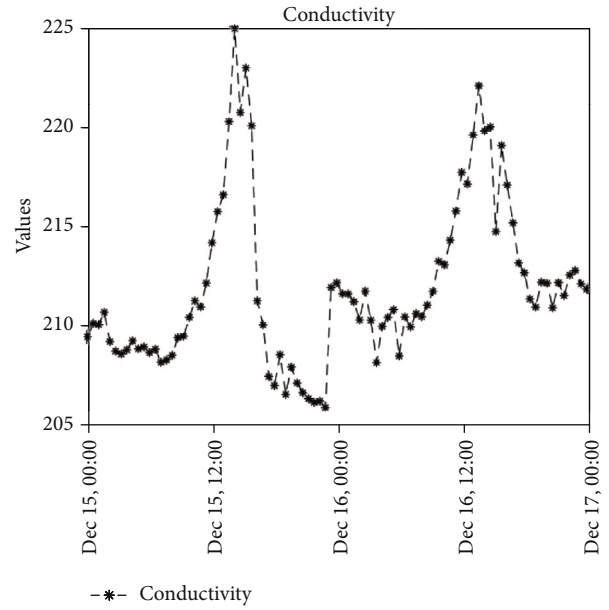


FIGURE 12: Readings from December 15 to 17, 2018.

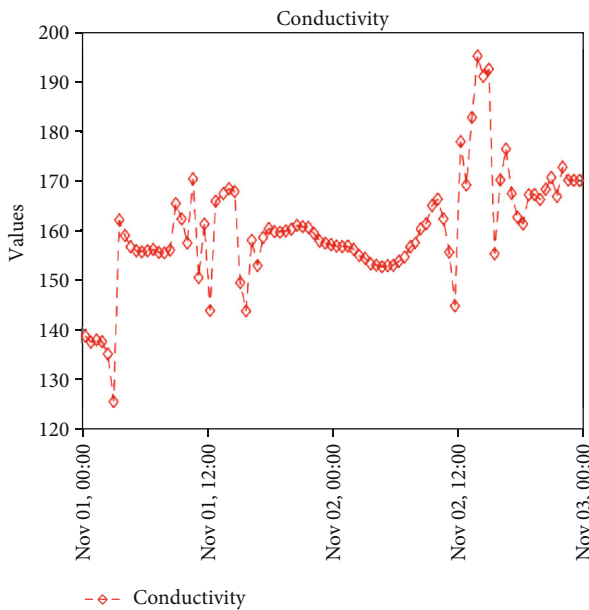


FIGURE 11: Readings from November 1 to 3, 2018.

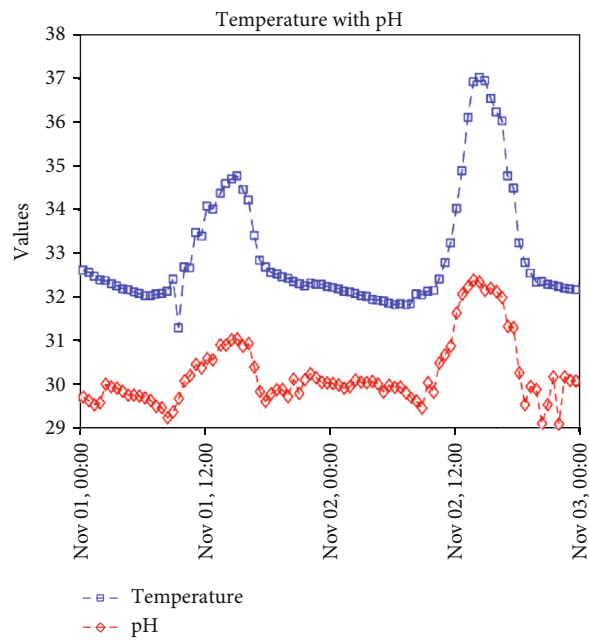


FIGURE 13: Temperature vs. pH readings from November 1 to 3, 2018.

4. WSN Design Experimentation, Results, and Discussions

This section presents the experimental results and discusses the results.

4.1. Experimental Results. During the period of deployment, the data collected were observed over time. We noticed several challenges during the data collection and transfer phase: (1) equipment malfunctioning; (2) damage of water quality sensor probes; (3) untimely recharge of internet data by the service provider and interruptions in data transmission; and (4) data-value out-of-range. The challenges encountered in the data collection and transfer phase

required researchers to perform data validation and reliability analysis. Data validation was performed to ascertain the quality and reliability of the data since errors will affect stakeholder's decision and planning. Gaps and inconsistencies were observed in some of the values reported by some of the sensor probes.

4.2. Discussions. In this section, discussions and conclusions are provided. Figures 9-32 show the graphs obtained from the sensor readings for some of the days within the period of deployment. Although the deployment area is a reserved

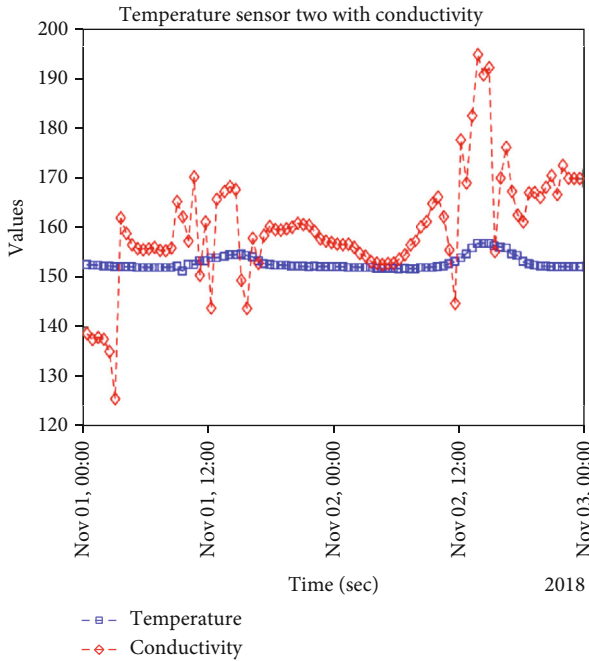


FIGURE 14: Temperature vs. conductivity readings of November 1-3, 2018.

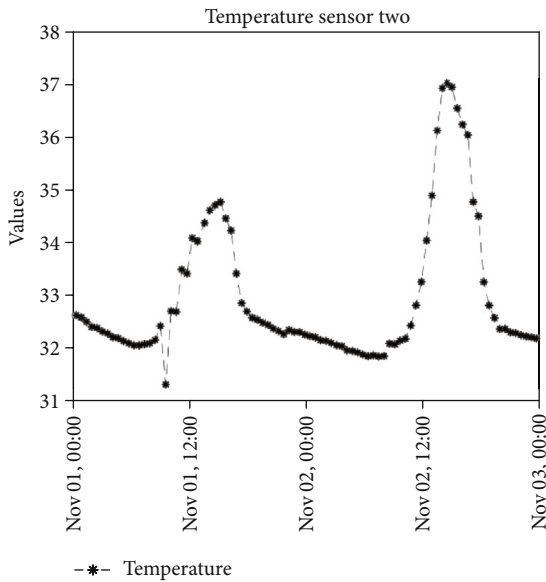


FIGURE 15: Temperature sensor two readings from November 06 to 08, 2018.

area, and activities such as fishing and farming were restricted, some fishing activities were illegally seen at certain times on the river. These fishermen used the freshwater weeds to trap fishes, and after their operation, they leave the weeds to float downstream.

The weeds cover sections (Figure 33) of the river resulting in habitat alteration and depletion of oxygen in the river. Apart from freshwater weeds, the presence of harmful algal blooms (HABs) was detected after a month of deployment especially when the dry season period was ushering in from

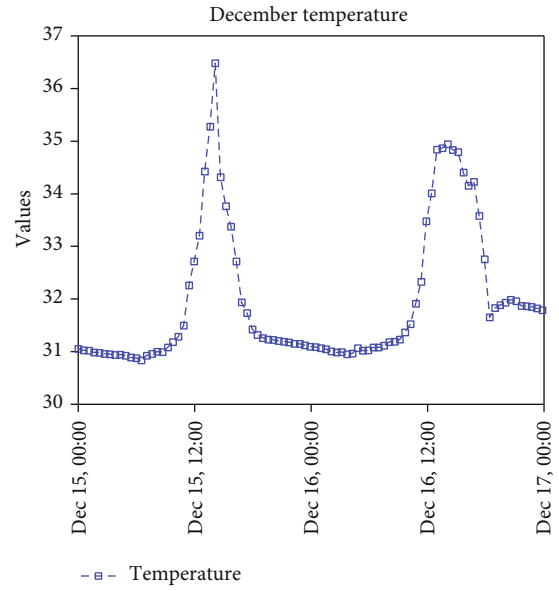


FIGURE 16: Temperature sensor two readings of December 15-17, 2018.

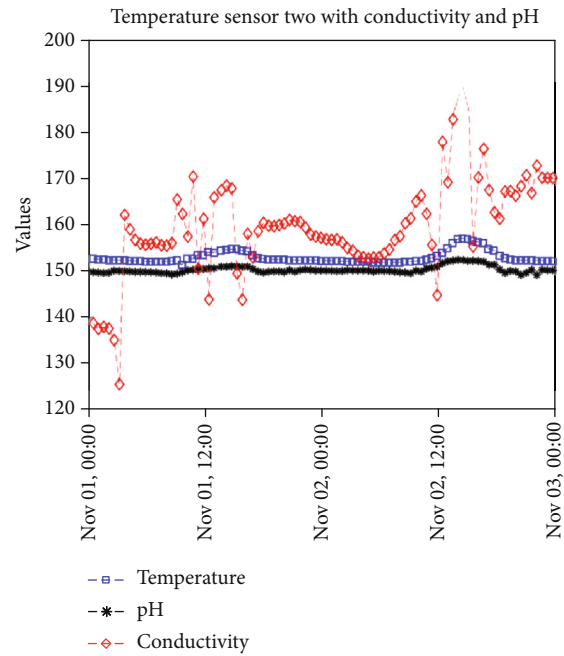


FIGURE 17: Temperature vs. pH and conductivity readings from November 1 to 3, 2018.

November 16, 2018, to January 28, 2019 (see Figure 34). Between the two months, we observed that the oxygen content was low due probably to the presence of HABs and freshwater weeds. This had a severe effect on aquatic life. HAB domination over a long period of time was also due to the activities happening upstream. Farming activities along the river stretch also saw the presence of pesticides and other harmful chemicals detrimental to aquatic life flow gradually into the river. The results obtained from this study indicate how WSNs may be used to predict activities happening upstream enabling stakeholders to use these data to develop

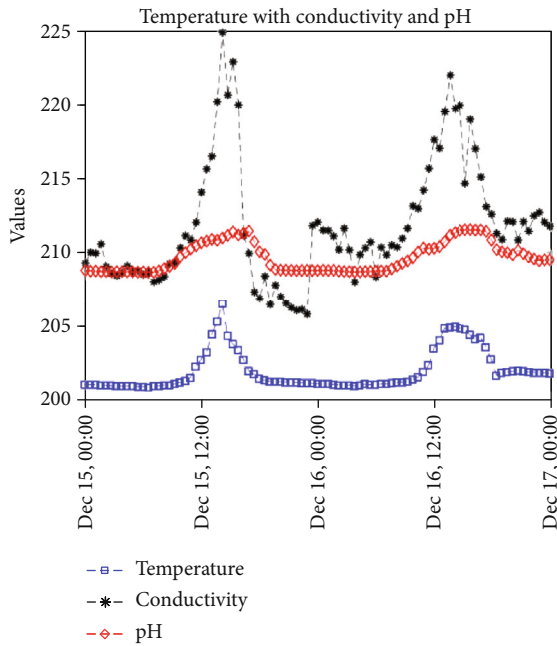


FIGURE 18: Temperature vs. pH and conductivity readings from December 6 to 8, 2018.

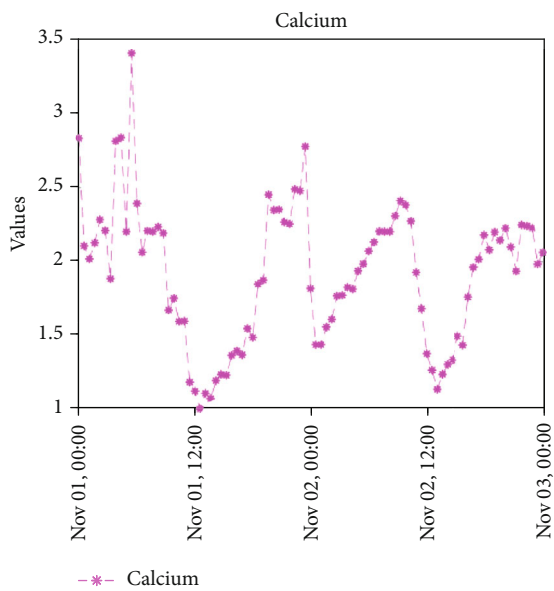


FIGURE 19: Calcium sensor readings from November 1 to 3, 2018.

and design strategies and come out with environmental management programs and other interventions for the community, region, and the nation as a whole for individuals to exert positive attitudes towards our freshwater resources.

The results presented are based on the data obtained from three Waspnote sensor devices as shown in Table 5. pH may be defined as the logarithmic concentration of hydrogen (H+) and hydroxide (OH-) ions of a substance constituting $H + +OH - = H2O$. The pH values obtained ranged between 6.75 and 14, as can be observed from the graphs shown in Figures 9 and 10. It shows the variations in pH in the river at different times of the day. The pH values

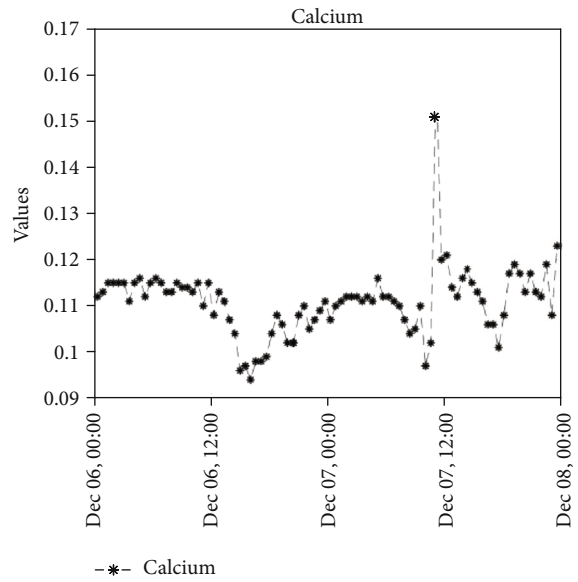


FIGURE 20: Calcium sensor readings from December 6 to 8, 2018.

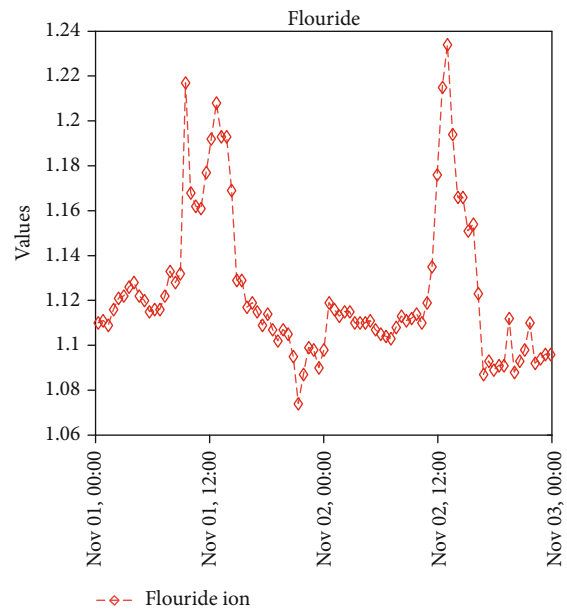


FIGURE 21: Fluoride sensor readings from November 1 to 3, 2018.

recorded above the pH value of 7 indicates that the Weija intake is more basic or alkaline and therefore supports aquatic animals much better.

The level of alkalinity of a river may be attributed to the location of the river as well as activities happening upstream. Other atmospheric factors contributed to the increase of pH over time, especially in the month of November 2018. Electrical conductivity is the ability of a medium to carry electrical energy. The electrical conductivity of water is directly proportional to the number of salt ions in the water. The temperature of the river, the geology of the area, climatic changes (e.g., from rainy to sunny seasons), and pollution from human waste or some industrial activities affects the conductivity levels in freshwater bodies.

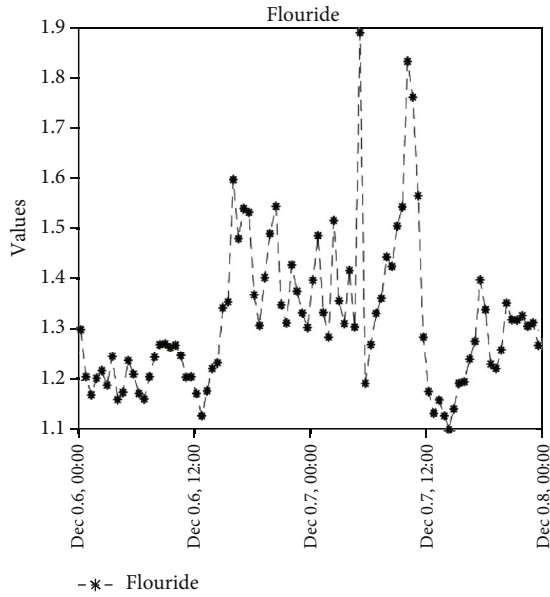


FIGURE 22: Fluoride sensor readings from December 6 to 8, 2018.

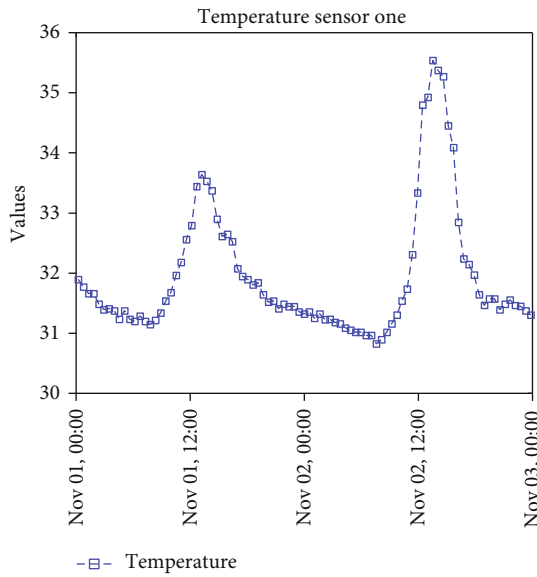


FIGURE 23: Temperature sensor one readings from November 1 to 3, 2018.

Conductivity values ranging from 0 to 200micro S/cm are considered as low conductivity. Conductivity values ranging from 200 to 1000micro S/cm are considered within the mid-range of conductivity levels, and values ranging between 1000 to 10000micro S/cm are considered as high conductivity values. The higher the river's temperature, the higher the conductivity value. The conductivity sensors were calibrated at a room temperature of 30°C. Aquatic animals who live in freshwater sources do not require high levels of conductivity. High levels of conductivity indicated in Figure 11 and Figure 12 in November and December may also have led to the death of some aquatic animal such as fishes as shown in Figures 35 and 36.

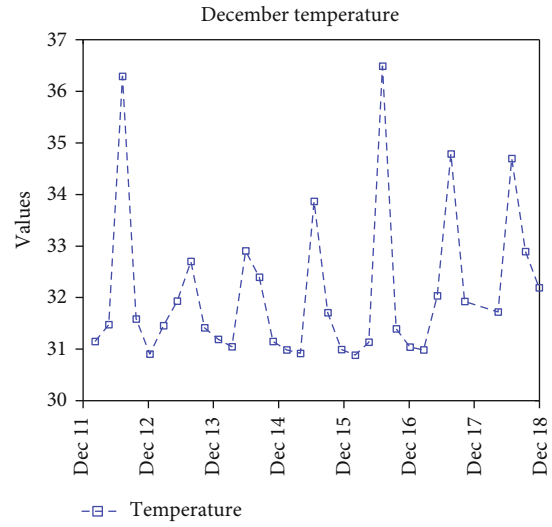


FIGURE 24: Temperature sensor one readings of December 11-18, 2018.

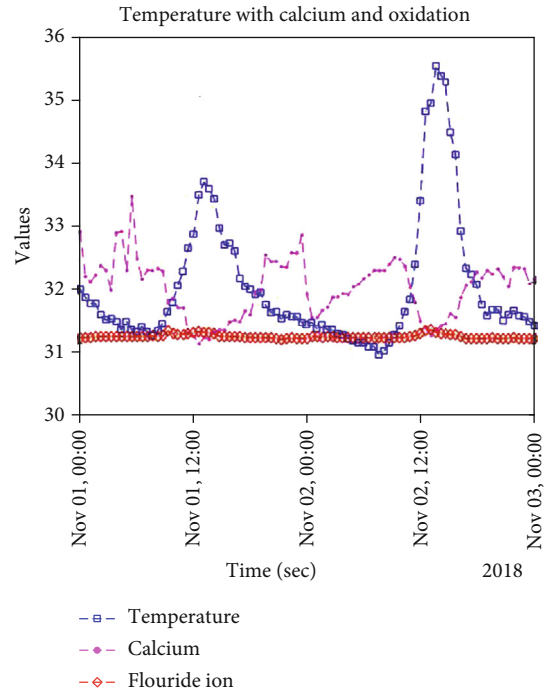


FIGURE 25: Temperature vs. calcium and fluoride readings from November 1 to 3, 2018.

Figure 31 shows that the temperature of the river rose to 37°C (also see Figures 15 and 16). This rise in temperatures was a result of the high levels of algae bloom in the river around that period, and that maybe the possible cause of conductivity levels rising to 196S/cm in November 2018 (see Figure 11) to 225 S/cm in December 2018 as shown in Figure 12. In Figures 13 and 14, we plot the temperature readings, pH, and conductivity over November and December 2018 to show the periods in which the rise in temperature had impacted the levels of conductivity in the river under consideration (also see Figures 17 and 18). The pH values

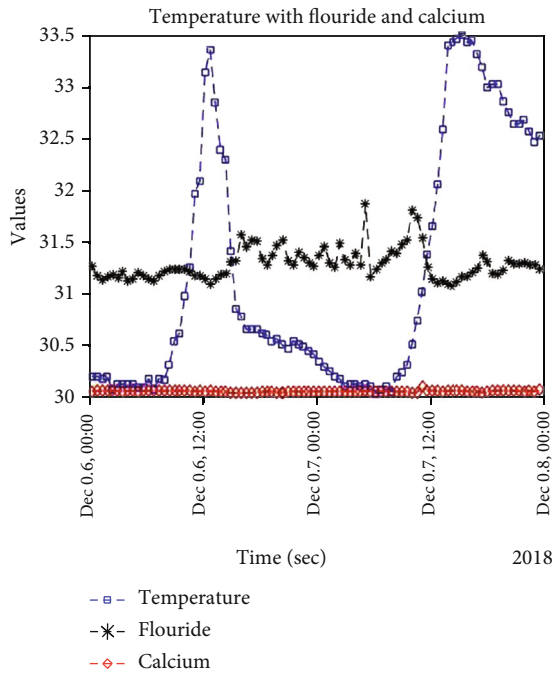


FIGURE 26: Temperature vs. calcium and fluoride readings from December 6 to 8, 2018.

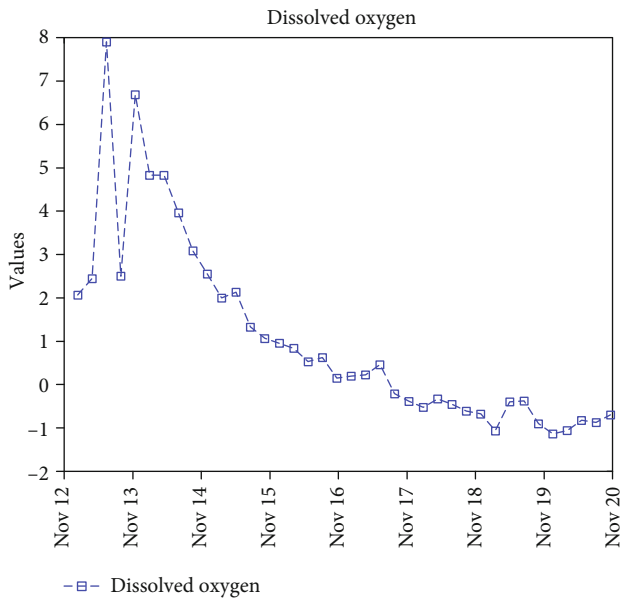


FIGURE 27: Dissolved oxygen sensor readings from November 12 to 20, 2018.

were scaled up by 150, and the temperature values were also scaled by 120 to obtain the plot shown in Figure 13. Although several factors affect the increasing levels of conductivity in river bodies, a study conducted by [36], revealed an essential property of rivers which is self-purification from pollution. Therefore, we observed that some points in December 2018, especially December 16, 2018, conductivity levels dropped to 205 S/cm, which may be attributed to the self-purification process probably around that time.

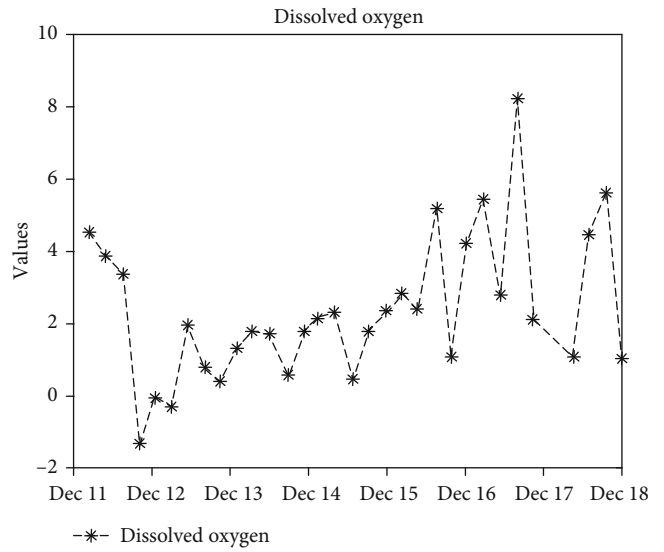


FIGURE 28: Dissolved oxygen sensor readings from November 11 to 18, 2018.

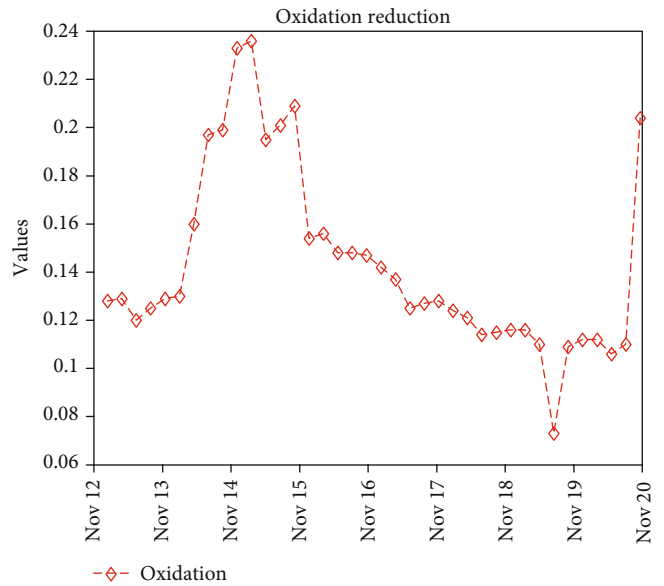


FIGURE 29: Oxidation sensor readings from November 12 to 20, 2018.

The geological structure (such as the type of soil, plantation, and weather conditions) at the Weija intake contributed to the presence of calcium in the river [37]. The study showed that calcium levels rose to about 3.5 mg/L in the month of November 2018 and dropped to about 0.16 mg/L in December 2018, as shown in Figure 20. The temperature sensor connected to this Waspnote sensor unit measured temperature of the river up to about 35°C in November 2018 and in December 2018 temperatures rose to about 36°C, as shown in Figure 23. Fluoride levels measured in November 2018 rose to about 1.24 mg/L, and in December 2018, the amount of fluoride in the river rose to about 1.9 mg/L, as shown in Figure 7. In Figures 25 and 26, the calcium, fluoride, and temperature readings from November to December 2018 are

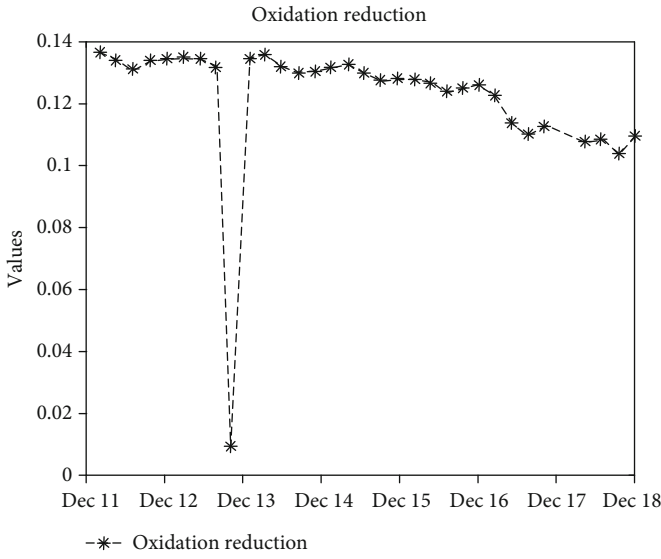


FIGURE 30: Oxidation sensor readings from December 11 to 18, 2018.

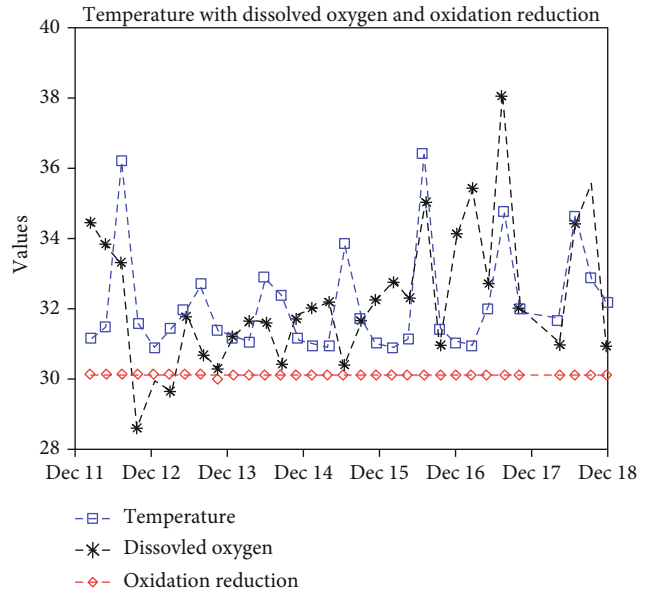


FIGURE 32: Temperature vs. dissolved oxygen and ORP readings of December 11-18, 2018.

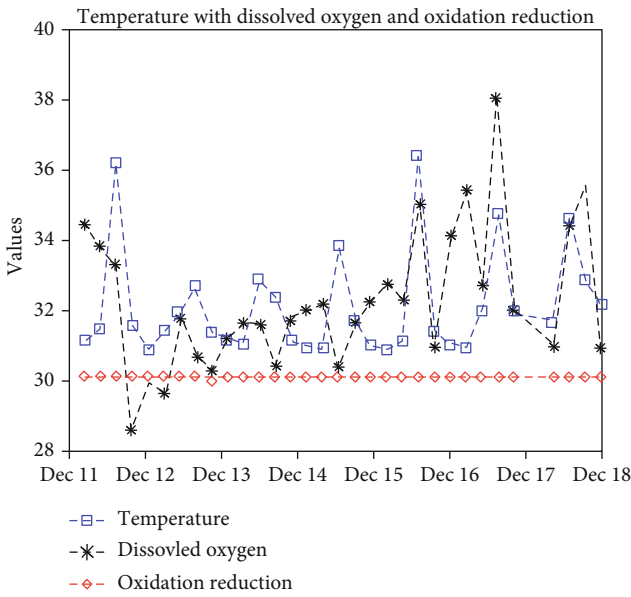


FIGURE 31: Oxidation sensor readings from November 12 to 20, 2018.

presented. Figures 29 and 30 show the ORP graphs. The ORP value declined due to the lower amount of oxygen content in that time of day (see Figures 31 and 6).

The amount of DO in a river determines the number and type of organisms that live in the river [38]. Fish, for example, depends on the amount of dissolved oxygen in the river body. The study revealed high levels of DO content in the river from October to November 13, 2018 (see Figures 27 and 28). These DO levels significantly dropped from November 13 to negative DO values. This was because of the high levels of algae bloom and other grasses that covered the surface of the river rendering less oxygen penetration into the river. This it is believed might have attributed to the death of aquatic species in the river (see Figures 35 and 36). However,



FIGURE 33: Weeds cover sections of river.



FIGURE 34: HABs in the river.

in December, the oxygen content rose up from the negative DO to reach 8 mg/L. Temperatures recorded by the sensor connected to this Waspnote unit indicated that most of the



FIGURE 35: Dead fishes in the river.



FIGURE 36: Low DO content causing aquatic fishes to die.

time in November and December 2018, the river was warm recording temperatures between 32 and 37°C (Figures 7 and 32). Warmer temperatures decrease the oxygen content in the river. This also affects aquatic life [39].

5. Future Directions

This paper provides several vital opportunities to explore deeper into significant issues related to sensor-based water quality monitoring. For example, to conserve the energy of the sensor nodes, researchers in the future should be looking into designing models that can predict water quality data based on historical data to regulate the sensor's energy consumption. Researchers in the future should also be interested in investigating the design of persuasive technology models (i.e., applying physiological principles to persuade people), data analysis models, and sensors that are capable of measuring both physical and chemical parameters. Monitoring water quality in real time in developing countries is vital, and this paper provides the initial steps.

5.1. Designing Prediction Models. Due to the energy consumption of nodes during data communication and the necessity to have continuous data during water quality monitoring, predictive models have been suggested. Predictive models reduce data communication while still producing

highly accurate data. Water parameters that are monitored could produce either linear or nonlinear data. Linear data could be predicted using time series models, while nonlinear data could be predicted using neural networks or support vector machines [40]. Water quality data are usually a combination of both linear and nonlinear. Predicting combined linear data is more comfortable to implement with low complexities but are limited in predicting the nonlinear variations in the data. Examples of linear models include autoregressive (AR), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and Grey Series Models.

ARMA models are for predicting slow-changing data while MA models are for data with sudden and sharp changes. Nonlinear models are somewhat challenging to implement but give better accuracy. Examples include neural networks, decision trees and rule learners, and genetic algorithms. The limitation of neural networks in WSNs is that while the data sets are mostly centralised, making them not suitable for distributed ADHOC deployments, they also require a longer time to train the models. Several WQM approaches use a combination of some linear and nonlinear models [41, 42]. Much work must be done to efficiently use time series models for near accurate predictions that could be used effectively in solving problems related to ADHOC wireless networks such as mobility and topology changes, energy limitations, routing, and ADHOC deployment.

5.2. Designing of Persuasive Technology Models. Water quality monitoring could be a useful tool in influencing the behaviour of people living at the banks of river bodies who use the water directly without further treatment. Persuasive technologies are a general class of technologies to apply physiological principles of persuasion such as credibility, trust, reciprocity, and authority to change intended users' attitudes and behaviours [43]. Persuasion is with the intentional effort to change attitudes and behaviours using technologies from abstracted ideas such as water quality monitoring. It would be imperative if applications were developed that inform an intended audience about the water quality parameters and give alerts that suggest the changes in these qualities, especially when they go beyond the acceptable thresholds. Such an application will not only change people's attitudes to the devices but also provide them with a sense of ownership and how they relate towards the water and the environment at large.

6. Conclusion

This paper focused on a real-time deployment of a wireless sensor network application for collecting water quality measurements for five months, September 2018 up to January 2019. In this project, we used smart sensor nodes from Libelium to measure the water quality parameters. The implementation provides a sustainable approach that monitors freshwater sources to support citizens in a particular locality where there is no access to clean potable water but rely on lakes, streams, rivers, and boreholes, enabling them to know the characteristics of the water they are drinking. The measured data value is transferred to cloud storage for

analysis. We designed a web portal using a set of protocols to process the captured data. Compared to other real-time deployed systems across the world, the solution presented here is secured, uses efficient MAC layer protocol which enhances the transmission and reception time and implements a network layer protocol that takes care of the delay factor in the transmission process. At the network level, we design the setup in such a way that when a parameter is measured, and the communication channel is not available, we perform retransmission until the communication channel is made available. Each of the parameters measured is as follows: calcium ion (Ca²⁺), electrical conductivity, pH, dissolved oxygen, fluoride ion (F), nitrate ion (NO₃⁻), oxidation-reduction potential (ORP), and temperature indicated significant fluctuations over time. These changes may be attributed to pollution from upstream, which are time varying. We plan to increase the number of sensor nodes taking into account sensor probes such as the turbidity sensor probe and total dissolved solid sensors since they are the parameters the stakeholders requested that we measure for them during the prototype phase.

Data Availability

The authors have data obtained for the experimentation but currently the data is not available for public use due to instructions from the Ghana Water Company whose permission under which the site used for the experiment was granted. Live data values are presented at the following web portal: <http://51.38.128.131/libelium/public/index.php>. Access codes may be granted if the Ghana Water Company grants us the permission to share this data.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank UG-Carnegie “Next Generation of Academics in Africa Project (BANGA Africa Corporation)”, the University of Ghana, for funding the project. We are grateful to them for the support provided to purchase the sensor nodes and other materials that made the implementation of this work successful.

References

- [1] D. Chalchisa, M. Megersa, and A. Beyene, “Assessment of the quality of drinking water in storage tanks and its implication on the safety of urban water supply in developing countries,” *Environmental Systems Research*, vol. 6, no. 1, p. 12, 2018.
- [2] K. Saravanan, E. Anusuya, R. Kumar, and L. H. Son, “Real-time water quality monitoring using internet of things in scada,” *Environmental monitoring and assessment*, vol. 190, no. 9, p. 556, 2018.
- [3] K. S. Adu-Manu, C. Tapparello, W. Heinzelman, F. A. Katsriku, and J.-D. Abdulai, “Water quality monitoring using wireless sensor networks: current trends and future research directions,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 1, p. 4, 2017.
- [4] F. A. Katsriku, M. Wilson, G. G. Yamoah, J. D. Abdulai, B. M. A. Rahman, and K. T. V. Grattan, “Framework for time relevant water monitoring system,” in *Computing in Research and Development in Africa*, pp. 3–19, Springer, 2015.
- [5] R. D. Robarts, S. J. Barker, and S. Evans, “Water quality monitoring and assessment: current status and future needs,” *Proceedings of Taal2007: The 12th World Lake Conference*, p. 175, 2008.
- [6] P. Mohammadi, S. Lotfi, S. P. Moussavi, M. Mousazadeh, and R. Rostami, “Studying quality of drinking water and determining sustainable indicators for water resources of villages of Harsin town of Iran,” *International Journal of Health and Life Sciences*, vol. In Press, no. In Press, 2018.
- [7] M. Mirzabeygi, M. Yousefi, H. Soleimani, A. A. Mohammadi, A. H. Mahvi, and A. Abbasnia, “The concentration data of fluoride and health risk assessment in drinking water in the Ardakan city of Yazd province, Iran,” *Data in brief*, vol. 18, pp. 40–46, 2018.
- [8] D.-N. Le, R. Kumar, and J. M. Chetterjee, *Introductory concepts of wireless sensor network. theory and applications*, 2018.
- [9] Z. Rasin and M. R. Abdullah, “Water quality monitoring system using Zigbee based wireless sensor network,” *International Journal of Engineering & Technology*, vol. 9, no. 10, pp. 24–28, 2009.
- [10] C. Borden and D. Roy, *Water Quality Monitoring System Design*, 2019, February 2019, <https://www.iisd.org/sites/default/files/publications/water-quality-monitoring-system-design.pdf>.
- [11] M. Wu, L. Tan, and N. Xiong, “Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications,” *Information Sciences*, vol. 329, pp. 800–818, 2016.
- [12] M. T. Lazarescu, “Design and field test of a WSN platform prototype for long-term environmental monitoring,” *Sensors*, vol. 15, no. 4, pp. 9481–9518, 2015.
- [13] G. Xu, W. Shen, and X. Wang, “Applications of wireless sensor networks in marine environment monitoring: a survey,” *Sensors*, vol. 14, no. 9, pp. 16932–16954, 2014.
- [14] M. Shirode, M. Adaling, J. Biradar, and T. Mate, *Iot Based Water Quality Monitoring System*, 2018.
- [15] W.-Y. Chung and J.-H. Yoo, “Remote water quality monitoring in wide area,” *Sensors and Actuators B: Chemical*, vol. 217, pp. 51–57, 2015.
- [16] S. Sridharan, “Water quality monitoring system using wireless sensor network,” *International Journal of Electronic Communications Engineering Advanced Research*, vol. 3, pp. 399–402, 2014.
- [17] B. Guanochanga, R. Cachipuendo, W. Fuertes et al., “Real-time air pollution monitoring systems using wireless sensor networks connected in a cloud-computing, wrapped up web services,” in *Proceedings of the Future Technologies Conference*, pp. 171–184, Springer, 2018.
- [18] P. Loreti, A. Catini, M. De Luca, L. Bracciale, G. Gentile, and C. Di Natale, “The design of an energy harvesting wireless sensor node for tracking pink iguanas,” *Sensors*, vol. 19, no. 5, p. 985, 2019.
- [19] I. E. Radoj, J. Mann, and D. K. Arvind, “Tracking and monitoring horses in the wild using wireless sensor networks,” in *2015 IEEE 11th International Conference on Wireless and Mobile*

- Computing, Networking and Communications (WiMob)*, pp. 732–739, Abu Dhabi, United Arab Emirates, October 2015.
- [20] T. Le Dinh, W. Hu, P. Sikka, P. Corke, L. Overs, and S. Brosnan, “Design and deployment of a remote robust sensor network: experiences from an outdoor water quality monitoring network,” in *32nd IEEE Conference on Local Computer Networks (LCN 2007)*, pp. 799–806, Dublin, Ireland, October 2007.
- [21] F. Regan, B. O’Flynn, A. Lawlor, J. Wallace, J. Torres-Sanchez, and S. C. Ó. Mathúna, “Experiences and recommendations in deploying a real-time, water quality monitoring system,” *Measurement Science and Technology*, vol. 21, no. 12, p. 124004, 2010.
- [22] Y. Luo, K. Yang, Z. Yu et al., “Dynamic monitoring and prediction of Dianchi Lake cyanobacteria outbreaks in the context of rapid urbanization,” *Environmental Science and Pollution Research*, vol. 24, no. 6, pp. 5335–5348, 2017.
- [23] A. N. Prasad, K. A. Mamun, F. R. Islam, and H. Haqva, “Smart water quality monitoring system,” in *2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, pp. 1–6, Nadi, Fiji, December 2015.
- [24] A. M. C. Ilie, C. Vaccaro, J. Rogeiro, T. E. Leitão, and T. Martins, “Configuration, programming and implementation of 3 smart water network wireless sensor nodes for assessing the water quality,” in *2017 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/IUC/ATC/CBDCCom/IOP/SCI)*, pp. 1–8, San Francisco, CA, USA, August 2017.
- [25] F. Ahmedi, L. Ahmedi, B. O’Flynn et al., “In WaterSense: an intelligent wireless sensor network for monitoring surface water quality to a river in kosovo,” in *Innovations and Trends in Environmental and Agricultural Informatics*, pp. 58–85, IGI Global, 2018.
- [26] N. Karimi, A. Arabhosseini, M. Karimi, and M. H. Kianmehr, “Web-based monitoring system using wireless sensor networks for traditional vineyards and grape drying buildings,” *Computers and Electronics in Agriculture*, vol. 144, pp. 269–283, 2018.
- [27] K. S. Adu-Manu, N. Adam, C. Tapparello, H. Ayatollahi, and W. Heinzelman, “Energy-harvesting wireless sensor networks (EH-WSNS): a review,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 14, no. 2, p. 10, 2018.
- [28] D. J. Sullivan, J. K. Joiner, K. A. Caslow et al., *Continuous Monitoring for Nitrate in USGS Water Science Centers Across The US*, 2018, December 2018, <https://pubs.usgs.gov/of/2018/1059/ofr20181059.pdf>.
- [29] M. Koch, “Calibration,” in *Quality Assurance in Analytical Chemistry: Training and Teaching*, pp. 183–200, Springer, 2010.
- [30] “Data Frame Guide | Libelium,” 2019, September 2020, <http://www.libelium.com/development/waspmote/documentation/data-frame-guide/>.
- [31] A. Valada, C. Tomaszewski, B. Kannan, P. Velagapudi, G. Kantor, and P. Scerri, “An intelligent approach to hysteresis compensation while sampling using a fleet of autonomous watercraft,” in *International Conference on Intelligent Robotics and Applications*, pp. 472–485, Springer, 2012.
- [32] F. Regan, A. Lawlor, B. O. Flynn et al., “A demonstration of wireless sensing for long term monitoring of water quality,” in *2009 IEEE 34th Conference on Local Computer Networks pages*, pp. 819–825, Zurich, Switzerland, October 2009.
- [33] K. S. Adu-Manu, F. Katsriku, J. D. Abdulai, J. M. Gómez, and W. Heinzelman, “Network Lifetime Maximization with Adjustable Node Transmission Range,” in *Smart Cities/Smart Regions—Technische, wirtschaftliche und gesellschaftliche Innovationen*, pp. 693–707, Springer Vieweg, Wiesbaden, 2019.
- [34] F. Engmann, F. A. Katsriku, J. D. Abdulai, K. S. Adu-Manu, and F. K. Banaseka, “Prolonging the lifetime of wireless sensor networks: A review of current techniques,” *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8035065, 2018.
- [35] F. K. Shaikh and S. Zeadally, “Energy harvesting in wireless sensor networks: A comprehensive review,” *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 1041–1054, 2016.
- [36] A. S. Adekunle and I. T. K. Eniola, “Impact of industrial effluents on quality of segment of Asa river within an industrial estate in Ilorin, Nigeria,” *New York Science Journal*, vol. 1, no. 1, pp. 17–21, 2008.
- [37] A. Potaszniak and S. Szymczyk, “Magnesium and calcium concentrations in the surface water and bottom deposits of a river-lake system,” *Journal of Elementology*, vol. 20, no. 3, 2015.
- [38] M. P. C. Agency, “Low Dissolved Oxygen in Water Causes, Impact on Aquatic Life? An Overview,” 2009.
- [39] “How exactly does dissolved oxygen affect water quality? | APEC Water,” *How Exactly Does Dissolved Oxygen Affect Water Quality?*, 2019, September 2020, https://www.freedrinkingwater.com/water_quality/quality1/1-how-dissolved-oxygen-affects-water-quality.htm.
- [40] F. Engmann, F. A. Katsriku, J.-D. Abdulai, and K. S. Adu-Manu, “Reducing the energy budget in WSN using time series models,” *Wireless Communications Mobile Computing*, vol. 2020, Article ID 8893064, p. 15, 2020.
- [41] D. Ö. Faruk, “A hybrid neural network and ARIMA model for water quality time series prediction,” *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010.
- [42] G. Tan, J. Yan, C. Gao, and S. Yang, “Prediction of water quality time series data based on least squares support vector machine,” *Procedia Engineering*, vol. 31, pp. 1194–1199, 2012.
- [43] B. J. Fogg, “Persuasive technology: using computers to change what we think and do,” *Ubiquity*, vol. 2002, 2002.

Research Article

Censoring-Based Relay Transmission for Achieving Energy Efficiency in Cognitive Radio Networks

Li Feng ^{1,2}, Jirong Sun,¹ Hua Sun,¹ and Shengju Tang¹

¹School of Engineering & Technology, Sichuan Open University, Chengdu 610073, China

²Educational Information Management and Information System Research Center of Open University of China (OUC), Sichuan Open University, Chengdu 610073, China

Correspondence should be addressed to Li Feng; feng_fengli@163.com

Received 22 February 2020; Revised 31 July 2020; Accepted 6 August 2020; Published 8 September 2020

Academic Editor: Yujin Lim

Copyright © 2020 Li Feng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Energy efficiency (EE) is critical to achieve cooperative sensing and transmission in relay-assisted cognitive radio networks (CRNs) with limited battery capacity. This paper proposes an energy-efficient cooperative transmission strategy with combined censoring report and spatial diversity in cooperative process, namely, censoring-based relay transmission (CRT). Specifically, secondary relays (SRs) take part in cooperative sensing with differential censoring to reduce energy consumption, and the best SR assists secondary transmission to enhance communication quality in transmission stage and thus to improve secondary transmission EE. First, we derived generalized-form expressions for detection probability, reporting probability, sensing energy, and expected throughput for CRT. Second, we investigate a mean EE-oriented maximization nonconvex problem by joint optimizing sensing duration and power allocation for secondary users under secondary outage probability and sensing performance constraints. With the aid of Jensen's inequality, an efficient cross-iteration algorithm with low complexity is proposed to obtain the suboptimal solutions, which is developed by golden segmentation search method. Finally, extensive simulations are conducted to evaluate the performance of CRT. The results show significant improvements of SUs' EE compared with traditional noncooperation single cognitive transmission schemes, which demonstrate the benefits of our proposed cooperative strategy in conserving energy for secondary transmission.

1. Introduction

The energy consumption of communication systems has increased dramatically, making it an urgent issue in the information world to improve energy efficiency (EE) of mobile communications. The high energy consumption and exponential growth in wireless communication networks face serious challenges to the design of more EE and spectrum efficiency (SE) green communications that should deal with the scarcity of radio resources. A promising approach technique called cognitive radio networks (CRNs) is proposed as a key design to improve spectral efficiency of wireless communication. Especially, advancements in miniaturization of devices with higher computational capabilities and ultralow power communication technologies are driving forces for the ever growing deployment of embedded devices in our

surroundings; how to achieve energy-efficient transmission becomes even more critical [1–3]. Cognitive relay networks have received a lot of interest to improve network performance by expanding the coverage and increasing SE [4]. It is well known that cooperative relaying promotes a diversity gain, improving spectrum efficiency and reducing the time used to deliver a message through distributed transmission and signal processing [5, 6]. At the same time, the flexibility of cognitive radios is important to address the challenges and trade-offs between EE performance and practicality in [2, 7]. As important means for improving SE, cognitive radio and cooperative transmission are also considered potential techniques to achieve green communication. By introducing cooperative transmission into CRNs, cooperative CRNs can not only improve SE but also overcome fading effect and thus attract much attention from researchers [4–10]. In this

context, cooperation and cognition techniques can bring invaluable contributions to development of energy efficient networks.

In short, in this paper, our goal is to improve network EE while ensuring performance of secondary transmitting system. We propose an innovative best-relay-assisted transmission with censoring scheme to provide energy-efficient cooperative transmission in overlay CRNs. Our main contributions of this paper are summarized as follows:

- (i) We propose a novel energy-efficient and best-relay-assisted cooperative transmission scheme called CRT to reduce sensing energy overhead on sensing stage and enhance communication quality in transmission stage. CRT naturally integrates the censoring report, best-relay-assisted transmission, and time and power optimization in cooperative process, which can effectively improve secondary transmission EE
- (ii) On the basis of CRT scheme, we derive the generalized-form expressions for reporting probability, average energy, and average throughput for CRT. We formulate a mean EE-oriented maximization nonconvex problem by joint optimizing sensing duration and power allocation for secondary users under constraints of minimal sensing performance and secondary outage probability. Theoretical analysis demonstrates that the EE function of CRT scheme has a globally unique optimal solution. And then, an efficient cross-iteration algorithm was proposed to obtain the suboptimal solutions
- (iii) Finally, we conduct extensive simulations to evaluate EE performance of CRT which is also compared with that of the noncooperative scheme. It is shown that CRT scheme can remarkably improve EE metrics of secondary transmission compared to the non-cooperative case

The rest of this paper is structured as follows. Section 2 briefly describes any related work. The system model and transmission description is described in Section 3. In Section 4, the proposed optimization problem is formulated, and sensing time and power allocation strategy and the algorithm are presented in Section 5. Simulation results are discussed in Section 6. Finally, the conclusions are drawn in Section 7.

2. Related Work

As discussed above, fast increasing popularity of powerful intelligent devices and vast penetration of mobile internet business cause explosive volume of wireless traffic, which in turn makes spectrum even more crowded and communication systems more energy-intensive. Therefore, EE has been considered one of the key features in the 5th-generation mobile communication and has attracted many research interests [1, 4, 6]. As important means for improving SE, cognitive radio and cooperative transmission are also considered potential techniques to achieve green communication. By

introducing cooperative transmission into cognitive radio networks, cooperative CRNs can not only improve SE but also overcome fading effect and thus attract much attention from researchers.

Most recent research works on cooperative CRNs are mainly focused on improvement of access opportunities and transmission throughput; few of them are concerning on EE. However, it is necessary to study impacts on cooperative spectrum sensing (CSS) and cooperative transmission in cooperative CRNs when EE is concerned to meet low-carbon communication requirement. Given the importance of this issue, recent studies have focused on energy-efficient secondary transmission in CRNs from different aspects and indicated that cognitive and cooperation technology can reduce network energy consumption in CRNs [8–18]. Collaborative sensing also known as CSS is an efficient spectrum sensing technique to improve the sensing accuracy in cognitive radio. However, it brings extra collaborative sensing overhead due to mutual exchange of large information among cognitive users. In CSS, the number of cooperative users, fusion rule, transmission power, and sensing time affects the energy efficiency (EE) of the CSS. The authors in [8, 9], investigate subcarrier transmission path selection and power allocation optimization problem with the goal of maximizing system EE in cognitive relaying links considering secondary QoS and total power budget, and a step-by-step strategy based on binary search-assisted ascent method is proposed. In [11], the authors proposed a cluster-based CSS strategy to maximize EE by optimizing the fusion rule, whereby the transmission power and sensing time were presented by the joint optimization problem. In [13], a spectrum sharing strategy is proposed via multiwinner auction with multiple bands to increase throughput of secondary system under constraint of minimum EE requirement for primary system. In [15], an EE-oriented “win-win” cognitive spectrum sharing scheme is proposed in heterogeneous cognitive radio networks. The authors in [5] investigated a trade-off between the secrecy throughput and EE in CRNs and proposed a cooperative spectrum sharing paradigm to improve both the secrecy throughput and the energy efficiency of primary users. Comprehensively considering sensing energy and detection performance, a differential reporting adaptive energy-efficient spectrum detection scheme was proposed [10], which can remarkably reduce sensing energy consumption with only slightly degrading detection performance. In [16], the authors modeled power allocation problem with the goal of maximizing the average EE of cognitive users in the fast-fading scene subject to a transmission interruption probability constraint and proposed an efficient power allocation strategy by using the fractional programming and Lagrange dual method. Similar to [16], the authors in [17] investigated EE maximization problem with sensing time and transmitting power as optimization variables. Different from noncooperative single cognitive transmission (NST) in [16–18] studied a weighted convex objective function of relay numbers with the consideration of global detection probability and bit error rate of cooperative transmission which is solved by applying numerical analysis to obtain the optimal relay numbers.

Based on recent research in energy-efficient cooperative sensing and transmission in CRNs, the motivation of this paper is expressed as follows. Compared with single cognitive transmission [16, 17, 19], CSS improves spectrum detection accuracy of a cognitive radio network. Clearly, more secondary relays (SRs) involving in spectrum detection require more energy consumption. On the other hand, when secondary system needs to meet higher rate requirements, longer transmission distance will inevitably consume a large amount of transmitting power and appear to be inefficient. In [14], a combined censoring scheme is proposed with the goal of minimizing the network energy consumption subject to a specific detection performance constraint. It is shown that such a system can attain high energy savings. Notice that the essence of censoring mechanism is by decreasing the number of decision reporting information to reduce energy consumption on CSS. In other words, the result is only transmitted, if it is deemed to be informative. Moreover, [15, 18] also studied the maximization of energy-efficient transmission problem considering the energy consumption of sensing and transmission two stages, but they did not utilize censoring report. Besides, we investigate secondary transmission performance characterized by constraint of cognitive transmission outage probability, which differs from previous works [15–18] without QoS requirements.

Note that, in this paper, $\mathbb{E}[\cdot]$ denotes the expectation operator; $\Pr \{\cdot\}$ represents the probability of a random variable; $\Upsilon^{-1}(\cdot)$ denotes the inverse function of $\Upsilon(\cdot)$; $|\mathcal{E}|$ denotes the number of set \mathcal{E} ; $\mathcal{F}_{\mathcal{S}}$ represents the feasible region of programming; and \cdot stands for the definition operator.

3. System Model and Transmission Description

3.1. System Model. Consider a cognitive radio system with the coexistence of a pair of primary users (PU) $U_{P1} - U_{P0}$, a pair of secondary users ST-SD, and M secondary relay users SRs ($\mathcal{E} = \{\text{SR}_i | i = 1, 2, \dots, M\}$) that are available to assist secondary sensing and transmission [4, 17, 18]. Primary spectrum owned by the primary network is divided into several narrowbands with fixed bandwidth. Like [2, 17], without loss of generality, we assume that the channels are modeled as independent Rayleigh fading, where $h_{LI} (L \in H, I \in H, L \neq I, H = \text{ST}, \text{SD}, \text{SR}_i | i = 1, 2, \dots, M)$ denotes the fading coefficient. Besides, the additive white Gaussian noise (AWGN) received at $J = \{\text{ST}, \text{SRs}\}$ is denoted as n_j which is with zero mean and power spectral density σ_n^2 . From [16], when PU's signal is a complex PSK, we know that the false alarm probability and detection probability of energy detection (ED) with predefined detection threshold Δ_T for cognitive users are, respectively, written as $\phi_{f,k}(\tau) = Q(\kappa_1 \sqrt{\tau f_s})$ and $\phi_{d,k}(\tau) = Q(-\kappa_2 \sqrt{\tau f_s \kappa_3})$, where f_s is the sampling frequency, $\kappa_1 = \Delta_T / \sigma_n^2 - 1$, $\kappa_2 = 1 + \gamma_j - \Delta_T / \sigma_n^2$, $\kappa_3 = 1 / (2\gamma_j + 1)$, and $\Lambda(z) = (1 / \sqrt{2\pi}) \int_z^\infty e^{-u^2/2} du$. It is noted that, in this paper, the AND rule is used for decision fusion criterion to improve secondary system transmission opportunity as well as satisfy detection threshold Q_d^{th} . For implementation simplicity, we assume that the received signal-to-noise ratio γ_j at SU_j are the same

value, i.e., $\gamma_j = \gamma$. We let H_0 (i.e., $\theta = 0$) and H_1 (i.e., $\theta = 1$) denote two standard hypotheses used in spectrum sensing corresponding to PU's absence and presence, respectively. According to [13, 15], the global detection probability and false alarm probability of CSS based on AND fusion rule can be, respectively, written as

$$Q_{f,\text{AND}}(\tau) = \prod_{k \in J} \phi_{f,k}(\tau) = \left(\Lambda\left(\kappa_1 \sqrt{\tau f_s}\right) \right)^{M+1}, \quad (1)$$

$$Q_{d,\text{AND}}(\tau) = \prod_{k \in J} \phi_{d,k}(\tau) = \left(\Lambda\left(-\kappa_2 \sqrt{\tau f_s \kappa_3}\right) \right)^{M+1}. \quad (2)$$

Here, it should be pointed out that the secondary transmission opportunity is generally satisfied $Q_{f,\text{AND}}(\tau) \leq \phi_{f,k}(\tau) < 0.5$ in a cognitive radio network. So we can easily obtain $\kappa_1 > 0$ from $\phi_{f,k}(\tau)$. On the other hand, to avoid interference with PU's transmission, the global detection probability $Q_{d,\text{AND}}(\tau)$ needs to be not lower than a predetermined detection probability value Q_d^{th} , and then, we can obtain the relation $\phi_{d,k}(\tau) \geq Q_{d,\text{AND}}(\tau) \geq Q_d^{\text{th}} \geq 0.9 > 0.5$. Following, we have $\kappa_2 > 0$ by analyzing $\phi_{d,k}(\tau)$. To facilitate the upcoming analysis, the main parameters used are listed in Table 1 that are used in our analytical model. Based on the above discussion, we propose the CRT scheme to achieve energy-efficient transmission for overlay CRNs.

3.2. Scheme Description. As illustrated in Figure 1, CRT scheme adopts both censoring report and best-relay-assisted strategies, which is completely different from the traditional noncooperative single cognitive transmission (NST). Each media access control (MAC) frame of SU is composed of sensing phase duration τ and transmission phase of duration $T_S - \tau$ [5–10]. We assume that SRs operate in a fixed time division multiple access (TDMA) manner which is commonly considered in existing studies [16, 17], where each MAC frame is composed of two consecutive durations called spectrum sensing phase and data transmission phase. CRT is periodically executed both in sensing and transmission phase at the beginning of each MAC frame. Besides, the following transmission phase is the same as conventional packet transmission processes [5–17]. The communication process of CRT scheme is described in detail below.

- (i) Cooperative sensing with censoring: As can be seen from Figure 1, each sensing process with DCR consists of two essential parts: local decision phase and decision report phase occupying β and $1 - \beta$ fractions of τ , respectively. The decision report phase is further split $(1 - \beta)\tau$ into M equal segments. Then, each decision report phase is organized as M subslots $\{\tau_0, \tau_1, \dots, \tau_M\}$ occupying $(1 - \beta)/M$ fraction of τ . In DCR, ST holds a M bits buffer denoted as $\text{ST}_{\text{buf}} = [0, 0, \dots, 0]_{1 \times M}$ to store all SR_i 's latest reporting decision. Accordingly, SR_i needs a 1-bit buffer denoted as $\text{SR}_{\text{buf}}^k = [0]_{1 \times 1}$ to save previous local decision result. SR_1 makes its local decision $\hat{\theta}_{R1}$ first and sends $x_{R1} = 1$

only when $\widehat{\theta}_{R_i} \oplus \text{SR}_{\text{buf}}^i = 1$. Hence, differential censoring report strategy in CRT can be formulated as

$$\begin{cases} \text{SR}_i \text{ send } x_{\text{SR}_i} = 1, & \text{if } \widehat{\theta}_{R_i} \oplus \text{SR}_{\text{buf}}^i = 1, \\ \text{SR}_i \text{ keep quiet,} & \text{if } \widehat{\theta}_{R_i} \oplus \text{SR}_{\text{buf}}^i \neq 1, \end{cases} \quad (3)$$

where $x_{R_i} = 1$ means the current sensing result differs from its buffer content in τ_1 . Meanwhile, ST attempts to receive the indicator from SR_1 in τ_1 . In this case, if SR_1 's indicator is received successfully, ST inverses the bit $\text{SU}_{\text{buf}}(1)$ of buffer SU_{buf} and makes a decision by $\text{SU}_{\text{buf}}(1)$. Otherwise, ST makes a decision by $\text{SU}_{\text{buf}}(1)$ directly. Specific ideas of DCR strategy can be found in [5], which is omitted here due to space limitation

- (ii) Cooperative transmission by best-relay assisted: From Figure 1, transmission phase $T_S - \tau$ in CRT is further divided into two equal subslots, which are, respectively, denoted as T_{t1} and T_{t2} satisfying $T_{t1} = T_{t2} = 0.5(T_S - \tau)$. In T_{t1} , ST's data will broadcast to all secondary relay users SRs with transmit power P_s in the idle band. Then, SR_{k^*} with the largest end-to-end SNR is selected as best-relay-assisted user in cooperative secondary relay set $\Xi = \{\text{SR}_1, \dots, \text{SR}_p, \dots, \text{SR}_M\}$ and $|\Xi| = M$. The selection criterion of best-relay SUs can be described as $\text{SR}_{k^*} = \max_{\text{SR}_k \in \Xi} (\min(h_{\text{TR}_k}, h_{\text{R}_k\text{D}}))$. Motivated by the best-relay selection mechanism in [17], SR_{k^*} forward received data with transmitting power P_{R_k} to SD by decode and forward protocol in T_{t1} . It is worth mentioning that cooperative sensing and cooperative transmission stages are all assisted by secondary relays Ξ in CRT. In actuality, to save energy, ST can select an appropriate set of cooperative users based on instantaneous channel state [18]

4. Secondary Sensing and Transmission Optimization

4.1. Cooperative Sensing and Transmission Energy Overhead. Our main goal is to reduce the energy consumptions of sensing stage and satisfy sufficient accuracy. Next, we will examine the average energy of CSS, denoted as En_{CSS} . Without loss of generality, we assume that the circuit energy consumption of the secondary source users and the cooperative relay users are P_{ci}^S and P_{ci}^R and further assume that sensing and decision reporting consumption power for each cognitive user are denoted as P_{se} and P_{dr} [10]. Note that we ignore the constant of circuit consumptions in analysis of En_{CSS} ; i.e., En_{CSS} consists of P_{se} and P_{dr} only, due to the circuit power consumptions for each cognitive user which is always equal $\tau(P_{\text{ci}}^S + MP_{\text{ci}}^R)$ both in traditional and CRT schemes. By definition, false alarm means that PU is detected under H_0 and detection indicates that PU is detected under H_1 . Besides, miss detection is defined as that PU is undetected under H_1 . Like [5, 12, 15], as shown in Figure 2, whether PU is communicating or not can be modeled as a renewal process which

TABLE 1: System parameters.

Symbol	Description
h_{LI}	Channel coefficient between the link $L \rightarrow I$
T_S	Time slot duration
τ	Sensing phase duration of MAC frame
γ_j	Received signal-to-noise ratio at SR_j
β	Sensing time allocation factor
M	Number of available secondary relays
$P_{\text{ci}}^S, P_{\text{ci}}^R, P_S^{\text{max}}$	Secondary source, relay circuit power consumptions, and total transmission power
$\phi_{f,k}, \phi_{d,k}$	False alarm probability and detection probability of each SU
P_S, P_{R_k}	Transmitting power allocation on ST and SR_k
Φ	Transition matrix of PU's state
$Q_{f,\text{AND}}, Q_{d,\text{AND}}$	Global detection probability and false alarm probability of CSS
$P_{\text{se}}, P_{\text{dr}}$	Sensing, decision reporting consumption power for each SU
En_{CSS}	Average energy of CSS
$\text{Th}(\Theta_E)$	Secondary transmitting average throughput
$\text{En}(\Theta_E)$	Secondary transmitting average energy consumption
η_{EE}	Mean energy efficiency
ϕ_B, ϕ_I	Stationary probabilities for primary band to be busy and idle
ϕ_S^{CRT}	Secondary outage probability
Q_d^{th}	Minimum global detection probability of secondary transmission
ξ_S^{max}	Maximum outage probability of secondary transmission
π_1, π_0	Average busy and idle periods
σ_n^2	Noise power
$R_{LI}(x)$	Transmission data rate of the link $L \rightarrow I$

alternates between busy and idle states. The busy and idle periods can be assumed to be exponentially distributed which can be expressed as $f_B(t) = \nu e^{-\nu t}$ and $f_I(t) = \rho e^{-\rho t}$, where ν is the transition rate from busy to idle state and ρ is the transition rate from idle to busy state.

Accordingly, the average busy and idle periods are $\pi_1 = 1/\nu$ and $\pi_0 = 1/\rho$. Hence, the stationary probabilities for primary band to be busy and idle are given by $\phi_B = \pi_1/(\pi_1 + \pi_0)$ and $\phi_I = \pi_0/(\pi_1 + \pi_0)$. From [16], the transition matrix Φ is given by

$$\begin{aligned} \Phi &= \begin{bmatrix} \phi_{II}(t) & \phi_{BI}(t) \\ \phi_{IB}(t) & \phi_{BB}(t) \end{bmatrix} \\ &= \frac{1}{\pi_1 + \pi_0} \begin{bmatrix} \pi_0 + \pi_1 e^{-(\pi_0 + \pi_1)t} & \pi_1 - \pi_1 e^{-(\pi_0 + \pi_1)t} \\ \pi_0 - \pi_0 e^{-(\pi_0 + \pi_1)t} & \pi_1 + \pi_0 e^{-(\pi_0 + \pi_1)t} \end{bmatrix}. \end{aligned} \quad (4)$$

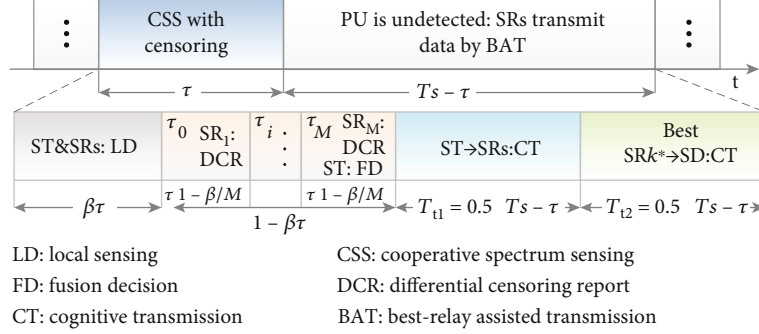


FIGURE 1: Time slot structure of proposed cooperative sensing and transmission.

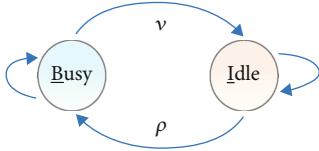


FIGURE 2: Continuous Markov transfer process of PU's channel.

For the purpose of comparison, the cooperative sensing energy consumption En_{CSS}^{tra} of traditional scheme without censoring can be derived as

$$En_{CSS}^{tra} = \tau \left[\beta(M+1)P_{se} + (1-\beta)P_{dr}\phi_{dr,R_i}^{Tra} \right], \quad (5)$$

where $\phi_{dr,R_i}^{Tra} = \phi_B(1-\phi_{d,R_i}) + \phi_I(1-\phi_{f,R_i})$ is the average decision reporting probability by a traditional dedicated reporting channel [17]. It should be mentioned that, in order to fully guarantee primary transmission QoS, secondary transmission opportunity is generally satisfied $\phi_{f,J} < 0.5 < \phi_{d,J}(\tau)$ in a cognitive radio network. Analyzing (5), we can get

$$En_{CSS}^{(\pi_1 < \pi_0)} > En_{CSS}^{(\pi_1 > \pi_0)}, \quad (6)$$

where $En_{CSS}^{(I)}$ indicates the cooperative sensing energy consumption when condition (I) in $En_{CSS}^{(I)}$ is satisfied.

Theorem 1. The average reporting probability by DCR is $\phi_{dr,R_i}^{CRT} = (o_1 - o_2)(\pi_0 + \pi_1)^{-2} e^{-(\pi_0 + \pi_1)t}$, where

$$\begin{cases} o_1 = \pi_0^2 \phi_{f,R_i} + \pi_0 \pi_1 (1 + \phi_{d,R_i} - \phi_{f,R_i}) + \pi_1^2 (1 - \phi_{d,R_i}), \\ o_2 = 2\pi_0 \pi_1 (\phi_{d,R_i} - \phi_{f,R_i}). \end{cases} \quad (7)$$

Proof. According to the initial status of PU, sensing result, and PU's transition cases in time slot τ_i , the SR decision censoring report scheme analysis for DCR can be, respectively, given as the following four possible scenarios as shown in Table 2.

TABLE 2: Four cases of SR_i decision by differential censoring report.

PU's initial state	PU's transition cases in τ_i	SR_i 's censoring report occurrence event
H_1	Busy \rightarrow busy	PU is undetected under H_1 (mis-detection)
	Busy \rightarrow idle	PU is undetected under H_0 (correct detection)
H_0	Idle \rightarrow idle	PU is detected under H_0 (false alarm)
	Idle \rightarrow busy	PU is detected under H_1 (detection)

From Table 2, decision censoring report probability ϕ_{dr,R_i}^{CRT} for SR_i in DCR can be calculated as

$$\begin{aligned} \phi_{dr,R_i}^{CRT} &= \sum_{n \in \{0,1\}} \sum_{X \in \{B,I\}} \phi(H_n) \Pr \left\{ \hat{\theta}_{R_i} = n | H_n \right\} \phi_{X|X}(t) \\ &= \phi(H_1) \phi_{B|B}(t) \Pr \left\{ \hat{\theta}_{R_i} = 0 | H_1 \right\} \\ &\quad + \phi(H_1) \phi_{I|B}(t) \Pr \left\{ \hat{\theta}_{R_i} = 0 | H_0 \right\} \\ &\quad + \phi(H_0) \phi_{I|I}(t) \Pr \left\{ \hat{\theta}_{R_i} = 1 | H_0 \right\} \\ &\quad + \phi(H_0) \phi_{B|I}(t) \Pr \left\{ \hat{\theta}_{R_i} = 1 | H_1 \right\} \\ &= \phi_B \left[\phi_{B|B}(t) (1 - \phi_{d,R_i}) + \phi_{I|B}(t) (1 - \phi_{f,R_i}) \right] \\ &\quad + \phi_I \left[\phi_{I|I}(t) \phi_{f,R_i} + \phi_{B|I}(t) \phi_{d,R_i} \right] \\ &= \phi_B + \phi_{d,R_i} \left(\phi_I \phi_{B|I}(t) - \phi_B \phi_{B|B}(t) \right) \\ &\quad + \phi_{f,R_i} \left(\phi_I \phi_{I|I}(t) - \phi_B \phi_{I|B}(t) \right). \end{aligned} \quad (8)$$

Therefore, substituting (4) into (8), after some algebra, ϕ_{dr,R_i}^{CRT} is proved.

In what follows, En_{CSS}^{CRT} represents the differential censoring sensing energy consumption in CRT, which can be calculated as

$$En_{CSS}^{CRT} = \tau \left[\beta(M+1)P_{se} + (1-\beta)P_{dr}\phi_{dr,R_i}^{CRT}(\tau(1-\beta)/M) \right]. \quad (9)$$

Next, we analyze the transmission rate and energy consumption for CRT scheme based on link status of $U_{P_1} - U_{P_0}$ and global decision of J . For ease of analysis, let Th_i ($i = 1, 2, 3, 4$) and En_i represent the secondary transmission rate and energy consumption in CRT. As shown in Figure 1, the throughput and power consumption in transmission stage can be discussed four scenarios as follows:

- (1) Presence detection with probability $\phi_B Q_{d,\text{AND}}(\tau)$: primary link is busy while sensing result is correct. In this case, in order to avoid interference to the primary system, the secondary system is prohibited transmission. Consequently, transmission rate is zero, and energy consumption in this case can be calculated as $\text{En}_1(\Theta_E) = \text{En}_{\text{CSS}}^{\text{CRT}} + T_S(P_{\text{ci}}^S + MP_{\text{ci}}^R)$, where $\Theta_E \triangleq [\tau, P_S, P_{R_k}]$
- (2) Miss detection with probability $\phi_B(1 - Q_{d,\text{AND}}(\tau))$: primary link is busy while sensing result is incorrect. In this case, like [12, 13], PU's signal usually has a large interference to secondary transmission, which causes SU to be unable to correctly decode data, so secondary transmission rate is zero and secondary energy consumption can be expressed as $\text{En}_2(\Theta_E) = \text{En}_{\text{CSS}}^{\text{CRT}} + T_S(P_{\text{ci}}^S + MP_{\text{ci}}^R) + 0.5(T_S - \tau)(P_S + P_{R_k})$
- (3) False alarm with probability $(1 - \phi_B)Q_{f,\text{AND}}(\tau)$: primary link is idle while sensing result is incorrect. There is no transmission with energy consumption $\text{En}_3(\Theta_E) = \text{En}_{\text{CSS}} + T_S(P_{\text{ci}}^S + MP_{\text{ci}}^R)$
- (4) Absence detection with probability $(1 - \phi_B)(1 - Q_{f,\text{AND}}(\tau))$: primary link is idle while sensing result is correct. Secondary transmission is successful with a rate of $0.5(T_S - \tau)\mathbb{E}[\widetilde{\text{Th}}]$, where $\widetilde{\text{Th}} = \min \{R_{TR_k}(P_S h_{TR_k}), R_{R_k,D}(P_{R_k} h_{R_k,D})\}$ and $R_{LI}(x)$ represents transmission data rate under bandwidth W , $R_{LI}(x) \triangleq W \log_2(1 + x/\sigma_n^2)$, $\mathbb{E}(\cdot)$ indicates the expectation operation, and secondary energy consumption in this case is calculated as $\text{En}_4(\Theta_E) = \text{En}_{\text{CSS}} + T_S(P_{\text{ci}}^S + MP_{\text{ci}}^R) + 0.5(T_S - \tau)(P_S + P_{R_k})$

Based on above four cases, the expected energy consumption for CRT can be expressed as

$$\begin{aligned} \overline{\text{En}}(\Theta_E) &= \sum_i \text{En}_i \Pr \{ \text{En} = \text{En}_i \} \\ &= \text{En}_{\text{CSS}}^{\text{CRT}} + T_S(P_{\text{ci}}^S + MP_{\text{ci}}^R) \\ &\quad + 0.5(T_S - \tau)(P_S + P_{R_k})\phi_A, \end{aligned} \quad (10)$$

where $\phi_A = 1 - \phi_B Q_{d,\text{AND}}(\tau) - \phi_I Q_{f,\text{AND}}(\tau)$. As discussed above, $Q_{d,\text{AND}}(\tau)$ usually needs to reach a predetermined detection value $Q_{d,\text{AND}}(\tau) \geq Q_d^{\text{th}} \geq 0.9$, and then, $\phi_B Q_{d,\text{AND}}(\tau) \approx \phi_B Q_d^{\text{th}}$. So, we can approximate $\overline{\text{En}}(\Theta_E)$ in [13] as follows:

$$\text{En}(\Theta_E) = \omega_1 + 0.5\omega_2\phi_A, \quad (11)$$

where $\omega_1 = \text{En}_{\text{CSS}} + T_S(P_{\text{ci}}^S + MP_{\text{ci}}^R)$, $\omega_2 = -(T_S - \tau)(P_S + P_{R_k})$, and $\phi_A = \phi_I Q_{f,\text{AND}}(\tau) + \phi_B Q_d^{\text{th}} - 1$.

Similarly, the expected secondary throughput in the CRT scheme can be derived as

$$\begin{aligned} \overline{\text{Th}}(\Theta_E) &= 0.5(T_S - \tau)\phi_I(1 - Q_{f,\text{AND}}(\tau))\mathbb{E} \\ &\quad \cdot [\min \{R_{TR_k}(P_S h_{TR_k}), R_{R_k,D}(P_{R_k} h_{R_k,D})\}]. \end{aligned} \quad (12)$$

As mentioned in Section 3.1, secondary link channel coefficient h_{IJ} obeys exponential distribution with mean λ_{IJ} , so it is impossible to obtain the closed expression $\mathbb{E}[\widetilde{\text{Th}}]$ in $\overline{\text{Th}}(\Theta_E)$. However, we observed that $R_{LI}(x)$ is a concave function. Therefore, $\widetilde{\text{Th}}$ is still a concave function due to the concavity of $\min(\log)$ function. In order to obtain the closed expression $\overline{\text{Th}}(\Theta_E)$, like [13], $\mathbb{E}[\widetilde{\text{Th}}]$ is approximated by Jensen inequality relationship, and it can be described as

$$\mathbb{E}[\widetilde{\text{Th}}] \leq \widehat{\text{Th}} = \min \{R_{TR_k}(P_S \lambda_{TR_k}), R_{R_k,D}(P_{R_k} \lambda_{R_k,D})\} \quad (13)$$

where $\widehat{\text{Th}} = \min \{R_{TR_k}(P_S \mathbb{E}[h_{TR_k}]), R_{R_k,D}(P_{R_k} \mathbb{E}[h_{R_k,D}])\}$. Furthermore, the expected secondary throughput in CRT can be expressed as

$$\text{Th}(\Theta_E) = 0.5(T_S - \tau)\phi_I(1 - Q_{f,\text{AND}}(\tau))\widehat{\text{Th}} \quad (14)$$

In summary, according to the primary link state and global decision probability, the secondary throughput and energy consumption expressions under a corresponding situation are discussed.

4.2. Energy-Efficient Resource Optimization Problem in CRT. Similar to [12, 13, 17], mean EE is defined as a ratio between average throughput and average energy consumption, i.e., $\eta_{\text{EE}}^{\text{CRT}}(\Theta_E) \triangleq \text{Th}(\Theta_E)/\text{En}(\Theta_E)$. Resource allocation in the CRT maximizes the mean EE as an optimization goal while considering secondary transmission QoS requirements. It is noted that the outage probability is introduced here to characterize secondary transmission QoS. According to the Shannon channel coding theorem, when channel capacity is lower than transmission rate required, it is considered a communication interruption. Thus, the secondary outage probability ϕ_S^{CRT} can be expressed as

$$\begin{aligned} \phi_S^{\text{CRT}} &= \Pr \left\{ 0.5(1 - \tau/T_S)\phi_I(1 - Q_{f,\text{AND}}(\tau))\widehat{\text{Th}} < \text{Th}_S^{\text{req}} \right\} \\ &= \phi_I(1 - Q_{f,\text{AND}}(\tau))\text{Pout}_S^{\text{CRT}}(\tau, P_S, P_{R_k}) \end{aligned} \quad (15)$$

where $\text{Pout}_S^{\text{CRT}}(\tau, P_S, P_{R_k}) \triangleq \Pr \{ 0.5(1 - \tau/T_S)\phi_I\widehat{\text{Th}} < \text{Th}_S^{\text{req}} \}$ and Th_S^{req} denotes the minimum secondary throughput. The EE maximization problem for CRT is denoted as P1, which is a resource allocation problem by joint optimizing sensing duration and power allocation for SU under constraints of sensing performance and minimal secondary outage probability. Mathematically, the optimization problem is modeled as

$$\begin{aligned}
\text{P1 : } \max_{\Theta_E} & \frac{0.5(T_S - \tau)\phi_I(1 - Q_{f,\text{AND}}(\tau))\mathbb{E}[\min\{R_{TR_k}(P_S h_{TR_k}), R_{R_k D}(P_{R_k} h_{R_k D})\}]}{\omega_1 + 0.5\omega_2\phi_A} \\
\text{s.t. } & \text{C1 : } Q_{d,\text{AND}}(\tau) \geq Q_d^{\text{th}}, \\
& \text{C2 : } \phi_B(1 - Q_{f,\text{AND}}(\tau))\text{Pout}_S^{\text{CRT}}(\tau, P_S, P_{R_k}) \leq \xi_S^{\text{max}}, \\
& \text{C3 : } P_S + P_{R_k} + P_{ci}^S + MP_{ci}^R + \text{En}_{\text{CSS}}^{\text{CRT}} \leq P_S^{\text{max}}, \\
& \text{C4 : } \beta^* = \arg_{\beta \in [0,1]} \min \text{En}_{\text{CSS}}^{\text{CRT}}(\tau), \\
& \text{C5 : } 0 \leq \tau \leq T_S, \quad P_S \geq 0, P_{R_k} \geq 0.
\end{aligned} \tag{16}$$

In P1, Q_d^{th} , ξ_S^{max} , and P_S^{max} , respectively, denote the minimum global detection probability, the maximum outage probability, and power consumption budgets of SUs. The constraint (16).C1 indicates that the global detection probability in CRT is not lower than a predefined detection threshold Q_d^{th} , thereby avoiding missed detection and causing interference to primary transmission. (16).C2 indicates that secondary outage probability is not greater than the present value ξ_S^{max} , and thus, secondary transmission quality is guaranteed. In addition, constraint relationship (16).C3 indicates that secondary total energy consumption needs to be no greater than total power consumption budgets of SUs P_S^{max} . It is noted that, like [5, 17], the saved sensing energy can be used for possible interweave transmission. Similar to the detection probability, $\text{En}_{\text{CSS}}^{\text{CRT}}$ can be minimized by optimizing β in CRT due to the fundamental tradeoff in sensing time allocation. So, it can be formulated as P2: $\min_{\beta} \text{En}_{\text{CSS}}^{\text{CRT}}(\tau)$ s.t.

$Q_{d,\text{AND}}(\tau) \geq Q_d^{\text{th}}$. Actually, minimum sensing energy can be achieved by optimizing β in CRT, which will be validated by the simulations in Section 6. In addition, both the objective function and constraints are very complicated, which makes solving P1 with a low complexity resource allocation strategy difficult. However, we observe that $\mathbb{E}[\widetilde{\text{Th}}]$ and $\widetilde{\text{Th}}$ in (13) have the same optimal sensing duration, which accounts for $\mathbb{E}[\widetilde{\text{Th}}]$ and $\widetilde{\text{Th}}$ that do not consist τ . Based on this analysis, the original optimization problem P1 can be decoupled into P3 for optimizing sensing time τ and P4 for optimizing power allocation $\{P_S, P_{R_k}\}$ separately. Specifically, secondary EE is only determined by sensing time τ at a fixed power $\{P_S^{\circ}, P_{R_k}^{\circ}\}$. Consequently, P3 is described as follows:

$$\begin{aligned}
\text{P3 : } \max_{\tau} \eta_{\alpha\text{EE}}^{\text{CRT}}(\tau) & = \frac{\text{Th}(\tau, P_S^{\circ}, P_{R_k}^{\circ})}{\text{En}(\tau, P_S^{\circ}, P_{R_k}^{\circ})} = \frac{\widetilde{\text{Th}}_{\alpha}(\tau)}{\widetilde{\text{En}}_{\alpha}(\tau)} \\
\text{s.t. } & \text{C1 : } \phi_B(1 - Q_{f,\text{AND}}(\tau))\text{Pout}_S^{\text{CRT}}(\tau, P_S^{\circ}, P_{R_k}^{\circ}) \leq \xi_S^{\text{max}}, \\
& \text{(16).C1, (16).C4, (16).C5,} \\
& \tag{17}
\end{aligned}$$

where $\text{Th}_{\alpha}(\tau) \triangleq \text{Th}(\tau, P_S^{\circ}, P_{R_k}^{\circ})$ and $\text{En}_{\alpha}(\tau) \triangleq \text{En}(\tau, P_S^{\circ}, P_{R_k}^{\circ})$ represent system throughput and energy consumption for a given transmission power $P_S^{\circ}, P_{R_k}^{\circ}$. Similarly, given a sensing time τ° , EE is only determined by SU's transmitting power in turn. Therefore, P4 can be expressed as

$$\begin{aligned}
\text{P4 : } \max_{P_S, P_{R_k}} \eta_{\beta\text{EE}}^{\text{CRT}}(P_S, P_{R_k}) & = \frac{\text{Th}(\tau^{\circ}, P_S, P_{R_k})}{\text{En}(\tau^{\circ}, P_S, P_{R_k})} = \frac{\text{Th}_{\beta}(P_S, P_{R_k})}{\text{En}_{\beta}(P_S, P_{R_k})} \\
\text{s.t. } & \text{C1 : } \phi_B(1 - Q_{f,\text{AND}}(\tau^{\circ}))\text{Pout}_S^{\text{CRT}}(\tau^{\circ}, P_S, P_{R_k}) \leq \xi_S^{\text{max}}, \\
& \text{C2 : } P_S + P_{R_k} \leq P_S^{\text{max}} - P_{ci}^S - MP_{ci}^R - \text{En}_{\text{CSS}}^{\text{CRT}} \text{ and (16).C5,} \\
& \tag{18}
\end{aligned}$$

where $\text{Th}_b(\tau) \triangleq \text{Th}(\tau^{\circ}, P_S, P_{R_k})$ and $\text{En}_b(\tau) \triangleq \text{En}(\tau^{\circ}, P_S, P_{R_k})$ denote the secondary throughput and energy consumption in CRT scheme for a given sensing time τ° , respectively.

5. Joint Sensing Time and Power Resource Allocation

In this section, a joint sensing time and power resource allocation for CRT is discussed. The following lemmas are provided to solve P3 and P4. Hereafter, notations $[F]_x' \triangleq \partial F(x)/\partial x$ and $[F]_x'' \triangleq \partial^2 F(x)/\partial x^2$ are defined the first derivative and the second derivative of F with respect to x for ease of presentation.

Theorem 2. *The objective function in P3 is a quasiconcave function with respect to τ .*

Proof. From (14), we can derive the first derivative of $\text{Th}_{\alpha}(\tau)$ as follows:

$$[\text{Th}_{\alpha}]_{\tau}' = 0.5\phi_B \left(Q_{f,\text{AND}}(\tau) - 1 + (\tau - T_S)(M + 1)\phi_f^M [\phi_f]_f' \right) \widehat{\text{Th}} \tag{19}$$

where $[\phi_f]'_\tau$ can be formulated as

$$[\phi_f]'_\tau = -\frac{\kappa_1}{2} \sqrt{\frac{f_s}{2\pi\tau}} e^{-(1/2)\tau f_s \kappa_1^2}, \quad (20)$$

and the second-order derivative of $\text{Th}_\alpha(\tau)$ can be further derived as

$$[\text{Th}_\alpha]''_\tau = \phi_B \left((M+1)\phi_f^M [\phi_f]'_\tau + 2(\tau - T_S) [Q_{f,\text{AND}}]''_\tau \right) \widehat{\text{Th}} \quad (21)$$

where $[Q_{f,\text{AND}}]''_\tau$ and $[\phi_f]''_\tau$ are, respectively, written as

$$[Q_{f,\text{AND}}]''_\tau = (M+1)\phi_f^{M-1} \left[M \left([\phi_f]'_\tau \right)^2 + \phi_f [\phi_f]''_\tau \right], \quad (22)$$

$$[\phi_f]''_\tau = \kappa_1 \left(\frac{1}{4\tau} + \frac{f_s \kappa_1^2}{4} \right) \sqrt{\frac{f_s}{2\pi\tau}} e^{-(1/2)\tau f_s \kappa_1^2}. \quad (23)$$

As discussion in Section 3.1, we have $\kappa_1 > 0$, $\kappa_2 > 0$, and $\sigma_n^2 < \Delta_T < \sigma_n^2(1 + \gamma)$. Analyzing (20) and (22), we have $[\phi_f]'_\tau < 0$ and $[\phi_f]''_\tau > 0$, which leads to $[Q_{f,\text{AND}}]'_\tau < 0$ and $[Q_{f,\text{AND}}]''_\tau > 0$, then $[\text{Th}_\alpha]''_\tau < 0$, which means $\text{Th}_\alpha(\tau)$ is concave function with respect to τ . Obviously, ω_1 in (11) is an affine function. It is not difficult to analyze that ω_2 in (11) is a convex function and monotonically increasing on τ , and ϕ_A is a convex function and monotonically decreasing on τ . From [19], $\omega_1 \omega_2$ is a convex function with respect to τ . Consequently, we can obtain that $\text{En}_\alpha(\tau)$ is convex function on τ . From $\text{Th}_\alpha(\tau)$ which is a concave function with respect to τ , we can ready analyze that $\eta_{\alpha\text{EE}}^{\text{CRT}}(\tau)$ is a quasiconcave function with respect to τ by the definition of quasiconcave function in [20] (see Appendix A for detail).

Afterward, analyzing constraint (17).C3, we can obtain the mathematical characteristics described in Theorem 2.

Theorem 3. *The constraint (17).C1 is equivalent to $\Omega(\tau) = \ln \omega_{S,1}^{\text{CRT}} + (\omega_{S,2}^{\text{CRT}} - 1)\omega_{S,3}^{\text{CRT}} \leq 0$ between $\tau \in [0, T_S]$, and $\Omega(\tau)$ is concave with respect to τ ; $\omega_{S,1}^{\text{CRT}}$, $\omega_{S,2}^{\text{CRT}}$, and $\omega_{S,3}^{\text{CRT}}$ in $\Omega(\tau)$ are, respectively, expressed as*

$$\begin{cases} \omega_{S,1}^{\text{CRT}} = (1 - \xi_S^{\max}) (\phi_I - Q_{f,\text{AND}}(\tau)\phi_I)^{-1}, \\ \omega_{S,2}^{\text{CRT}} = 2^{2R_s^{\text{req}} T_S / (W(T_S - \tau))}, \\ \omega_{S,3}^{\text{CRT}} = \sigma_n^2 \frac{(\lambda_{TR_k} P_S + \lambda_{R_k D} P_{R_k})}{(\lambda_{TR_k} \lambda_{R_k D} P_S P_{R_k})}. \end{cases} \quad (24)$$

Proof. See Appendix B.

According to Theorem 2 and Theorem 3, the optimal sensing duration time strategy is described in Theorem 4 specifically.

Theorem 4. *The optimal sensing time allocation strategy for P3 can be characterized as*

$$\tau^* = \max \left(\max \left(\tau^{\text{det}}, \underline{\tau} \right), \min \left(\tau^{\text{opt}}, \bar{\tau} \right) \right), \quad (25)$$

where $\tau^{\text{det}} = (f_s \kappa_3 \kappa_2^2)^{-1} (Q^{-1}(\sqrt{Q_d^{\text{th}}}))^2$, τ^{opt} , $\underline{\tau}$, and $\bar{\tau}$ are, respectively, written as

$$\begin{cases} \tau^{\text{opt}} = A^{-1}(0), & \tau \in [0, T_S], \\ \tau^{\text{Pout}} = B^{-1}(0), & \tau \in [0, T_S], \\ \underline{\tau} = \Omega^{-1}(0), & \tau \in [0, \tau^{\text{Pout}}], \\ \bar{\tau} = \Omega^{-1}(0), & \tau \in [0, \tau^{\text{Pout}}], \end{cases} \quad (26)$$

where $A(\tau) = [\eta_{\alpha\text{EE}}^{\text{CRT}} \Omega]'_\tau$, $B(\tau) = [\Omega]'_\tau$ and $Y^{-1}(\cdot)$ denotes the inverse function of $Y(\cdot)$.

Proof. The first derivative of $\phi_d(\tau)$ and $Q_{d,\text{AND}}(\tau)$ on τ are, respectively, derived as

$$[\phi_d]'_\tau = \frac{\Lambda_2}{2} \sqrt{\frac{\Lambda_3 f_s}{2\pi\tau}} \exp \left(-\frac{1}{2} \tau f_s \Lambda_3 \Lambda_2^2 \right), \quad (27)$$

$$[Q_{f,\text{AND}}]'_\tau = (M+1)\phi_f^M [\phi_d]'_\tau. \quad (28)$$

From (27), we can get $[\phi_d]'_\tau > 0$ and $\sigma_n^2 < \Delta_T < \sigma_n^2(1 + \gamma)$. Examining (28) leads to $[Q_{f,\text{AND}}]'_\tau > 0$, which indicates that $Q_{d,\text{AND}}(\tau)$ is a monotonically increasing function on τ , so constraint (16).C1 in P2 can be formulated as $\tau > \tau^{\text{det}} =$

$(f_s \Lambda_3 \Lambda_2^2)^{-1} (Q^{-1}(\sqrt{Q_d^{\text{th}}}))^2$. Additionally, $[\Omega]''_\tau > 0$ in Theorem 3 shows that $B(\tau)$ is a monotonically increasing function on τ . We have $B(\tau) \rightarrow -\infty$ as $\tau \rightarrow 0^+$ and $B(\tau) \rightarrow +\infty$ as $\tau \rightarrow T_S^-$. In other words, $\Omega(\tau)$ monotonically decreases with respect to τ in $[0, \tau^{\text{Pout}}]$ and monotonically increases with respect to τ in $[\tau^{\text{Pout}}, T_S]$. Therefore, the necessary and sufficient condition of (17).C1 existing a feasible τ in $[0, T_S]$ is $\Omega(\tau^{\text{Pout}}) \leq 0$. Furthermore, one of the following cases may occur for $\Omega(\tau^{\text{Pout}}) \leq 0$: (1) when $\tau \in [0, \tau^{\text{Pout}}]$, $\Omega(\tau)$ monotonically decreases on τ in $[0, \tau^{\text{Pout}}]$, and $\lim_{\tau \rightarrow 0^+} \Omega(\tau) > 0$, which indicates that there exists a unique $\underline{\tau}$ satisfying $\Omega(\underline{\tau}) = 0$; (2) when $\tau \in [\tau^{\text{Pout}}, T_S]$, similarly, we can get $\Omega(\tau)$ monotonically decreases on τ in $[\tau^{\text{Pout}}, T_S]$, and $\lim_{\tau \rightarrow T_S^-} \Omega(\tau) = +\infty$,

which indicates that there exists a unique $\bar{\tau}$. Conversely, if $\Omega(\tau^{\text{Pout}}) > 0$, there is no sensing time $\tau \in [0, T_S]$ to satisfy $\Omega(\tau) \leq 0$. Above all, the optimal sensing duration τ^* is obtained at the boundary point or the stagnation point $\tau^{\text{opt}} = A^{-1}(0)$. This depends on the zero point τ^{Pout} of the first derivative of (17).C1.

Similar to Theorem 2, $\eta_{\beta\text{EE}}^{\text{CRT}}(P_S, P_{R_k})$ of P4 can be proved a quasiconcave function with respect to $\{P_S, P_{R_k}\}$. We will

not go into details here to save space. Before solving P4, Theorem 5 is given.

Theorem 5. *In the CRT scheme, when secondary EE achieves maximum, the power allocation in P4 should be satisfied $\lambda_{TR_k} P_S = \lambda_{R_k D} P_{R_k}$.*

Proof. Prove by contradiction. According to the objective function in (16), secondary EE is not only related to ST \rightarrow SD but also limited by ST \rightarrow SR_k. Suppose that secondary system EE achieves optimum, there exists a $\{\tilde{P}_S, \tilde{P}_{R_k}\}$, satisfying $R_{TR_k}(\tilde{P}_S \mathbb{E}[h_{TR_k}]) \neq R_{R_k D}(\tilde{P}_{R_k} \mathbb{E}[h_{R_k D}])$. From (12), it can be seen that the secondary throughput $\text{Th}_b(\tilde{P}_S, \tilde{P}_{R_k}) = 0.5\hat{\text{Th}}$ indicates that its rate depends on the smaller rate, which means one can reduce the transmit power of the larger rate hop in $R_{TR_k}(\tilde{P}_S \mathbb{E}[h_{TR_k}])$ and $R_{R_k D}(\tilde{P}_{R_k} \mathbb{E}[h_{R_k D}])$, leading to $R_{TR_k}(\tilde{P}_S \mathbb{E}[h_{TR_k}]) = R_{R_k D}(\tilde{P}_{R_k} \mathbb{E}[h_{R_k D}])$. By the definition of EE function, i.e., $\eta_{EE}^{\text{CRT}}(\Theta_E) \triangleq \text{Th}(\Theta_E)/\text{En}(\Theta_E)$, reducing transmission power of the larger rate hop will increase system EE under the condition that the system rate is a constant. Obviously, the original hypothesis leads to contradiction. Therefore, the original proposition is true.

Theorem 5 shows that power allocation when secondary transmission EE achieves the maximum must satisfy the following relationship $P_{R_k}^* = P_S^* \lambda_{TR_k} \lambda_{R_k D}^{-1}$. So, the optimal power $\{P_S^*, P_{R_k}^*\}$ for P4 is described in Theorem 6 specifically.

Theorem 6. *The optimal power allocation strategy for P3 is*

$$\begin{cases} P_S^* = \max\left(\underline{P}_S, \min\left(K^{-1}(0), \overline{P}_S\right)\right), \\ P_{R_k}^* = P_S^* \lambda_{TR_k} \lambda_{R_k D}^{-1}, \\ \underline{P}_S = [2\sigma_n^2(\omega_{S,2}^{\text{CRT}} - 1)/(\lambda_{TR_k} \ln(1/\omega_{S,1}^{\text{CRT}}))]^+, \\ \overline{P}_S = (P_S^{\text{max}} - \text{En}_{\text{CSS}}^{\text{CRT}} - P_{ci}^S - MP_{ci}^R)/\left(1 + \lambda_{TR_k} \lambda_{R_k D}^{-1}\right), \end{cases} \quad (29)$$

where $[x]^+ \triangleq \max(x, 0)$ and $K(P_S) = [\eta_{\beta EE}^{\text{CRT}}]_P''$, $P_S \in [\underline{P}_S, \overline{P}_S]$.

Proof. According to (18).C2 in P4, the upper bound of P_S is obtained as

$$P_S \leq \overline{P}_S = \frac{P_S^{\text{max}} - P_{ci}^S - MP_{ci}^R - \text{En}_{\text{CSS}}^{\text{CRT}}}{1 + \lambda_{TR_k} \lambda_{R_k D}^{-1}}. \quad (30)$$

And from (18).C2 in P4, the lower bound of P4 can be derived as

$$P_S \geq \underline{P}_S = \frac{2\sigma_n^2(\omega_{S,2}^{\text{CRT}} - 1)}{\lambda_{TR_k} \ln(1/\omega_{S,1}^{\text{CRT}})}. \quad (31)$$

As discussed above, the optimal transmit power P_S^* for P4 can be obtained either at both ends of the boundary \underline{P}_S or \overline{P}_S taken at the stagnation point $P_S^{\text{opt}} = K^{-1}(0)$. This is depend-

ing on the constraints (18).C1 and (18).C2 in P4. In addition, when secondary system EE achieves maximum, each rate for two hops should be equal. It indicates that the optimal power allocation satisfies relationship $P_{R_k}^* \lambda_{R_k D} = P_S^* \lambda_{TR_k}$. So Theorem 6 is proved.

Theorem 6 shows that the optimal power allocation problem can be obtained according to (18) for a given sensing time τ° . In what follows, aiming to maximize secondary EE, a joint resource allocation algorithm for sensing time τ and transmission power $\{P_S, P_{R_k}\}$ on ST and SR_{k*} is proposed. We know that closed-form expressions of false alarm and detection probabilities for CRT are available in (1) and (2), which only require average channel gains and thus can be estimated in prior. Consequently, to solve P2, we develop linear search methods by golden section search (GSS) method [20]. The steps of optimal sensing time allocation factor β of P2 is expressed in Algorithm 1.

The pseudocode for CRT is presented in Algorithm 2. It is not difficult to observe that the key basis in Algorithm 2 is τ and $\{P_S, P_{R_k}\}$ alternating iterations according to Theorem 4 and Theorem 6 until $|ee^{[n+1]} - ee^{[n]}| < \xi_{ee}$ or does not satisfy the constraints of P3 and P4. For ease of analysis, we assume that sensing duration and transmission power after n iteration are denoted as $\{\tau^{[n]}, P_S^{[n]}, P_{R_k}^{[n]}\}$. Theorem 6 guarantees that $P_S^{[n]}, P_{R_k}^{[n]}$ is available during the n th iteration, and substituting $\tau^{[n]}$ into (29) has $\eta_{EE}^{\text{CRT}}(\tau^{[n]}, P_S^{[n+1]}, P_{R_k}^{[n+1]}) \geq \eta_{EE}^{\text{CRT}}(\tau^{[n]}, P_S^{[n]}, P_{R_k}^{[n]})$. Similarly, $\tau^{[n+1]}$ can be calculated by Theorem 4, that is, substituting $P_S^{[n+1]}, P_{R_k}^{[n+1]}$ into (25) and (26) has $\eta_{EE}^{\text{CRT}}(\tau^{[n+1]}, P_S^{[n+1]}, P_{R_k}^{[n+1]}) \geq \eta_{EE}^{\text{CRT}}(\tau^{[n]}, P_S^{[n]}, P_{R_k}^{[n]})$. In this way, the iterative process can continue until it is satisfied $|ee^{[n+1]} - ee^{[n]}| < \xi_{ee}$. Therefore, in Algorithm 2, the following relationship exists:

$$\eta_{EE}^{\text{CRT}}(\tau^{[n+1]}, P_S^{[n+1]}, P_{R_k}^{[n+1]}) \geq \eta_{EE}^{\text{CRT}}(\tau^{[n]}, P_S^{[n]}, P_{R_k}^{[n]}). \quad (32)$$

Namely, $\eta_{EE}^{\text{CRT}}(\tau^{[n]}, P_S^{[n]}, P_{R_k}^{[n]})(n = 1, 2, \dots)$ is increasing successively until convergence condition is satisfied.

Remark 7. Algorithm 2 for maximization secondary system EE of the CRT scheme has been proposed to obtain the optimal value Θ_E in (16). From (6), we observe that the system EE for the NST case with $M = 0$, which means the CRT scheme can degrade the NST scheme. Consequently, we can readily obtain the EE optimal resource allocation for the NST scheme by Algorithm 2, which will be validated by the simulations in Section 6.

6. Simulation Results and Discussions

In this section, we will evaluate the performance of proposed CRT by some numerical and simulation results and then also compare it with the traditional overlay case as presented in Section 3.1. Without loss of generality, the ED introduced in [16, 17] is used for local sensing throughout these


```

1: Initialize: Let  $\mu = (\sqrt{5}-1)/2$ ,  $\Gamma = \Gamma_0 = \{1, 2\}$ , tolerance  $\xi_\beta > 0$ 
   and  $[\beta^{lb}, \beta^{ub}]$  be the interval of GSS;
2: Compute  $\beta^1 = \beta^{lb} + (1-\mu)(\beta^{ub} - \beta^{lb})$ ;
3: Compute  $\beta^2 = \beta^{lb} + \mu(\beta^{ub} - \beta^{lb})$ ;
4: Repeat
5:     Find optimal sensing time allocation factor for a given  $\tau^{[n]}$ ;
6:     Calculate  $EN(n) = En_{CSS}^{CRT}(\tau^{[n]})$ ,  $(\beta_{n \in \Gamma} \in [0, 1])$ ;
7:     If  $En_{CSS}^{CRT}(\beta_1) \leq En_{CSS}^{CRT}(\beta_2)$  then
8:         Obtain  $\beta^{lb} = \beta_1$ ,  $\beta_1 = \beta_2$ ,  $\beta_2 = \beta^{lb} + \mu(\beta^{ub} - \beta^{lb})$ ;
9:         Calculate  $EN(1) = EN(2)$ ,  $\Gamma = \Gamma_0 \setminus \{2\}$ ;
10:    Else
11:        Obtain  $\beta^{ub} = \beta_2$ ,  $\beta_2 = \beta_1$ ,  $\beta_1 = \beta^{lb} + (1-\mu)(\beta^{ub} - \beta^{lb})$ ;
12:        Calculate  $EN(2) = EN(1)$ ,  $\Gamma = \Gamma_0 \setminus \{1\}$ ;
13:    Endif
14: Until  $|EN(2) - EN(1)| < \xi_\beta$  converges.
15: Output the optimal time allocation factor  $\beta^*$  of P2.

```

ALGORITHM 1: Obtain optimal sensing time allocation factor β for CRT.

```

1: Initialization Parameters: tolerance  $\xi_{ee} > 0$ , iteration index  $n = 0$ ;
2: Repeat
3:   Calculate  $\underline{P}_S = [2\sigma_n^2(\bar{\omega}_{S,2}^{CRT} - 1)/(\lambda_{TR_k} \ln(1/\bar{\omega}_{S,1}^{CRT}))]^+$ ;
4:   Compute  $\bar{P}_S = (P_S^{\max} - En_{CSS}^{CRT} - P_{ci}^S - MP_{ci}^R)/(1 + \lambda_{TR_k} \lambda_{R_k D}^{-1})$ ;
5:   Calculate sensing time allocation factor  $\beta^*$  of P1 by Algorithm 1
6:   If  $\max\{\underline{P}_S, 0\} < \bar{P}_S$  then (Solve P4 by Theorem 6)
7:     Calculate  $P_S^{\text{opt}} = K^{-1}(0)$  in  $P_S = [0, \bar{P}_S]$ ;
8:     Obtain optimal  $P_S^{[n]} = \max(\underline{P}_S, P_S^{\text{opt}})$  under  $\tau^{[n]}$ ;
9:     Calculate  $P_{R_k}^{[n]} = P_S^{[n]} \lambda_{TR_k} \lambda_{R_k D}^{-1}$  in Theorem 5;
10:    If  $\Omega(\tau) < 0$  &&  $\tau \geq \tau^{\text{det}}$  then (Solve P3 by Theorem 4)
11:      Calculate  $\tau^{\text{det}} = (f_s \kappa_3 \kappa_2^2)^{-1} (Q^{-1}(\sqrt{M+1} \sqrt{Q_d^{th}}))^2$ ;
12:      Obtain  $\tau^{\text{opt}} = A^{-1}(0)$  and  $\tau^{\text{pout}} = B^{-1}(0)$  in  $\tau = [0, T_S]$ ;
13:      Obtain  $\underline{\tau} = \Omega^{-1}(0)$  in  $\tau = [0, \tau^{\text{pout}}]$ ;
14:      Obtain  $\bar{\tau} = \Omega^{-1}(0)$  in  $\tau = [0, \tau^{\text{pout}}]$ ;
15:      Compute  $\tau = \max(\max(\tau^{\text{det}}, \underline{\tau}), \min(\tau^{\text{opt}}, \bar{\tau}))$ ;
16:      Obtain  $ee^{[n]} = \eta_{EE}^{CRT}(\tau^{[n]}, P_S^{[n]}, P_{R_k}^{[n]})$ ;
17:    Endif
18:  Endif
19:  Update  $n = n + 1$ ;
20: Until  $|ee^{[n+1]} - ee^{[n]}| < \xi_{ee}$  ||  $n > \text{MaxIter}$ ;
21: Output  $\tau^*$ ,  $P_S^*$ ,  $P_{R_k}^*$  and optimal EE.

```

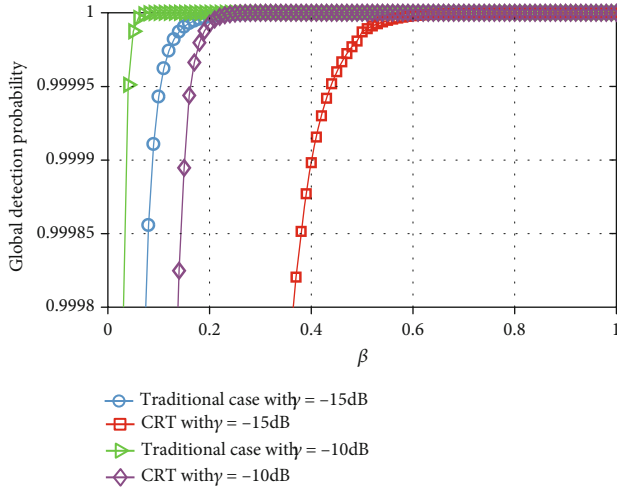
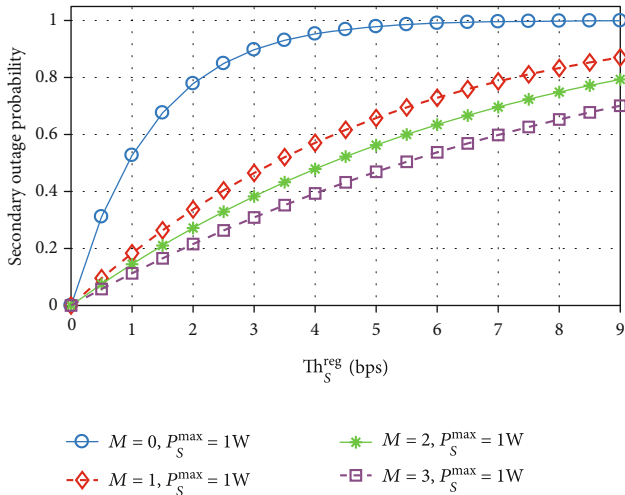
ALGORITHM 2: Joint sensing duration and transmission power allocation for CRT.

examples. Since we want to show the advantages of the proposed scheme, the choice of detector is not critical. Thus, the results obtained in this paper can be easily extended to other detector cases. Consider the scenario as follows. The secondary link distance between ST-SD is chosen as 1000 m, and SRs are randomly distributed in the interval [400, 600]m. Throughout these simulation studies, unless otherwise noted, the simulation parameters are provided in Table 3.

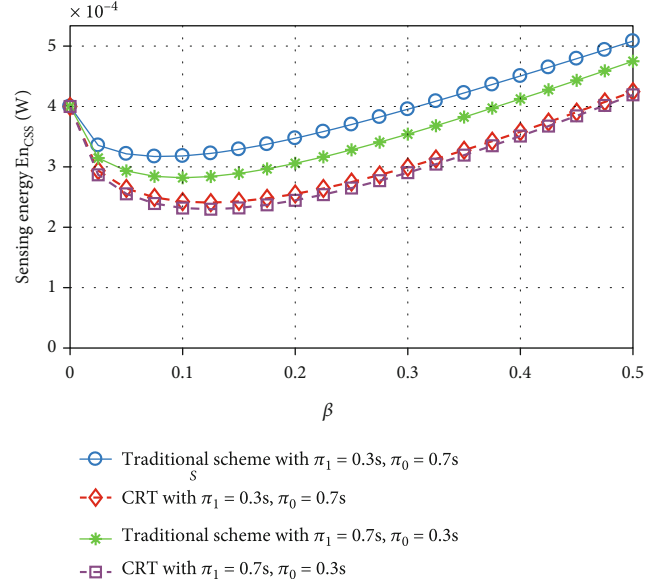
First, we plot the global detection probability $Q_{d,AND}$ versus the sensing time allocation factor β for the traditional and CRT schemes in Figure 3. One can observe from Figure 3 that CRT experiences performance degradation compared with the traditional scheme due to the less local sensing time and decision reporting time. However, the minimum performance gap is rather small. Besides, such performance gap becomes smaller as γ increases. Thus, under a reasonable detection probability requirement, CRT can not only

TABLE 3: Simulation experiment parameter values.

Parameters	Value	Parameters	Value
T_S	0.1 s	P_S^{\max}	10 W
π_1	0.3 s	f_s	10^6 Hz
π_0	0.7 s	R_S^{req}	0.5 bps
P_{ci}^R	0.02 W	P_{ci}^S	0.02 W
P_{se}	0.02 W	P_{dr}	0.04 W
σ_n^2	-64 dBm	Path loss exponent	3.5


 FIGURE 3: Q_d versus β for the traditional and proposed schemes.

 FIGURE 4: Secondary outage probability versus Th_S^{req} .

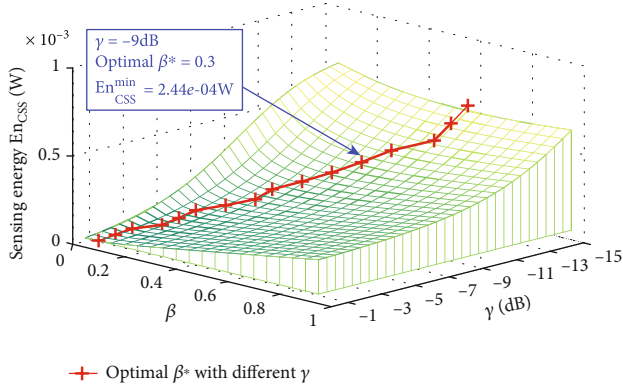
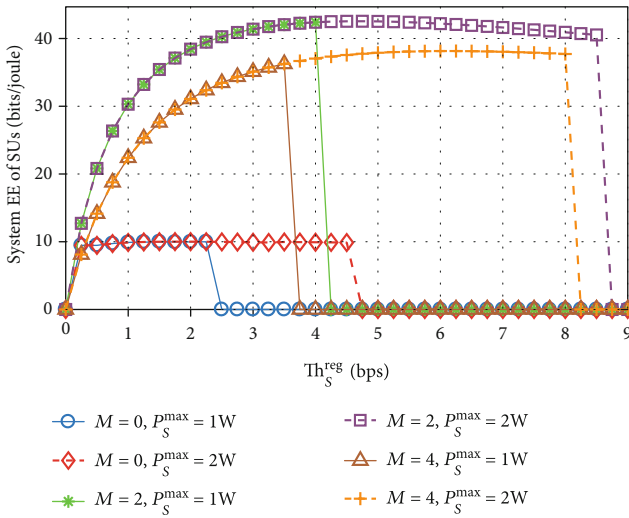
maintain the sensing accuracy but also cut the sensing energy overhead. This will be validated later. Clearly, the global detection probability $Q_{d,\text{AND}}$ increases with γ_p growing. We also see that β can be optimized to minimize the miss detection probability for CRT. Figure 4 shows the relationship between outage probability versus Th_S^{req} when different cooperative numbers are given. We observe that the best SR assists secondary communication to enhance QoS in the transmis-


 FIGURE 5: En_{CSS} versus β for the traditional and proposed schemes with $\gamma = -15$ dB.

sion stage. Furthermore, we also find that the secondary outage probability can be minimized by selecting a specific power allocation.

Second, we plot En_{CSS} versus β at $\gamma = -15$ dB for the traditional and CRT schemes under PU's behavior. As shown in Figure 5, it is observed that, under a given β , CRT remarkably reduce sensing energy consumption compared to the traditional case. It can be also noticed that En_{CSS} of the traditional scheme with the $\pi_1 = 0.3$ s and $\pi_0 = 0.7$ s case is larger than the $\pi_1 = 0.7$ s and $\pi_0 = 0.3$ s case, which accounts for the fact that SRs send less decision reporting information by adopting censoring report than that without dedicated reporting channel strategy [17], and confirms the conclusions made in Section 4.1. As a further development, Figure 6 depicts the three-dimensional plot between En_{CSS} and γ and β . It is also revealed that β can be optimized to minimize En_{CSS} to achieve minimizing the miss detection probability and sensing energy overhead for CRT, however, e.g., $\text{En}_{\min} = 2.44 \times 10^{-4}$ W and optimal sensing time allocation factor $\beta = 0.3$ under $\gamma = -9$ dB, which confirms the conclusions in Remark 7. Furthermore, we found that the optimal β^* will decrease as γ increases, for the reason that increasing γ improves the local sensing reliability and thus, PU is detected to be more energy efficient.

Third, Figure 7 demonstrates the secondary EE and power consumption versus Th_S^{req} . It is observed that, under a given secondary throughput requirement, CRT has remarkably improving EE compared to the traditional case in Figure 7. Moreover, the secondary transmission powers both in the traditional and CRT schemes increase with Th_S^{req} growing. Specifically, we can see clearly that the transmitting power is always increased with Th_S^{req} growing until achieving the maximum power budgets of SU. However, EE curves increase with the minimum rate requirement, while decrease slightly in larger Th_S^{req} . This is due to the fact that

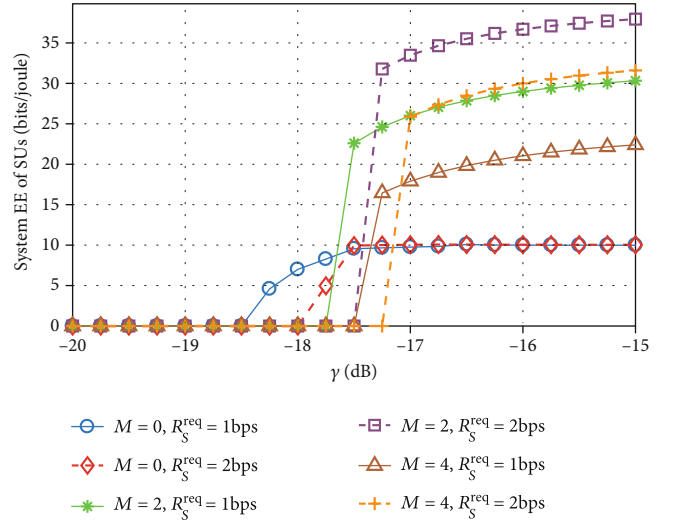
FIGURE 6: The three-dimensional plot between En_{CSS} and γ and β .FIGURE 7: Secondary system EE versus Th_S^{req} .

when the rate requirement is low, very low energy is needed to relay little data, which results in almost fixed throughput. On the other hand, when the rate requirement becomes high, more energy is needed to transmit data, which results in high expected throughput leading to a slight decrease in EE.

Finally, we depict the secondary EE versus γ in Figure 8. It is observed that CRT has significantly improving EE compared to the traditional case. Clearly, system EE is almost fixed and relatively small in the low γ region, which is due to the fact that secondary transmission is forbidden. However, as γ increases, system EE of CRT will increase, which accounts for the fact that the local sensing of SU will be sufficient to maintain detecting accuracy when γ is high.

7. Conclusion

In this study, we propose an innovative energy-efficient cooperative transmission strategy, namely, censoring-based relay transmission (CRT). CRT naturally integrates differential censoring report, best-relay-assisted transmission technologies in cooperative process, which can effectively improve secondary transmission EE by reducing energy on sensing stage and enhancing communication quality in the

FIGURE 8: Secondary system EE versus γ .

transmission stage. On the basis of CRT, we derive generalized-form expressions for reporting probability, average energy, and expected throughput for CRT. We investigate a mean EE-oriented maximization nonconvex problem by joint optimizing sensing duration and power allocation for secondary users under constraints of minimal sensing performance and secondary outage probability. An efficient cross-iteration algorithm with low complexity is proposed to obtain the suboptimal solutions, which is developed by the golden segmentation search method. Finally, simulation results show that the proposed cooperative paradigm is an efficient approach to boost both SUs' throughput and EE of secondary transmission remarkably compared to traditional noncooperation cognitive transmission schemes in high QoS requirements.

Appendix

A. Proof of Theorem 1

Proof. Let $\{\tau_1, \tau_2\} \in \mathcal{F}_S$, where \mathcal{F}_S represents the feasible region of P3 consisting the constraints (16).C1, (16).C5, and (17).C1, and further assuming that $0 \leq \vartheta \leq 1$ and $\eta_{\alpha EE}^{CRT}(\tau_1) < \eta_{\alpha EE}^{CRT}(\tau_2)$, since $Th_\alpha(\tau)$ is a concave function with respect to τ [10, 20], we can express the following concave function:

$$\begin{aligned} & Th_\alpha(\vartheta\tau_1 + (1-\vartheta)\tau_2) \\ & > \vartheta Th_\alpha(\tau_1) + (1-\vartheta)Th_\alpha(\tau_2) \\ & > \vartheta Th_\alpha(\tau_1) + (1-\vartheta)En_\alpha(\tau_2) \frac{Th_\alpha(\tau_1)}{En_\alpha(\tau_1)} \quad (A.1) \\ & = \frac{Th_\alpha(\tau_1)}{En_\alpha(\tau_1)} (\vartheta En_\alpha(\tau_1) + (1-\vartheta)En_\alpha(\tau_2)). \end{aligned}$$

In addition, $En_\alpha(\tau)$ is a convex function on τ , we have

$$En_\alpha(\vartheta\tau_1 + (1-\vartheta)\tau_2) < \vartheta En_\alpha(\tau_1) + (1-\vartheta)En_\alpha(\tau_2). \quad (A.2)$$

Substituting (A.2) into (A.1), the following relationship can be derived as

$$\frac{\text{Th}_\alpha(\vartheta\tau_1 + (1 - \vartheta)\tau_2)}{\text{En}_\alpha(\vartheta\tau_1 + (1 - \vartheta)\tau_2)} > \frac{\text{Th}_\alpha(\tau_1)}{\text{En}_\alpha(\tau_1)}. \quad (\text{A.3})$$

Analyzing (A.3), we can readily obtain that $\eta_{\alpha\text{EE}}^{\text{CRT}}(\tau)$ is a quasiconcave function with respect to τ by the definition of quasiconcave function in [20].

B. Proof of Theorem 2

Proof. This is due to h_{TR_k} and h_{R_kD} that obey exponential distribution with mean λ_{TR_k} and λ_{R_kD} , respectively. According to (15), we can derive the following:

$$\begin{aligned} (16).C2 &\Rightarrow \phi_I(1 - Q_{f,\text{AND}}(\tau)) \Pr \left\{ 0.5(1 - \tau/T_S) \widehat{\text{Th}} < \text{Th}_S^{\text{req}} \right\} \\ &\leq \xi_S^{\max} \Rightarrow \Pr \left\{ 0.5(1 - \tau/T_S) \widehat{\text{Th}} > \text{Th}_S^{\text{req}} \right\} \\ &\geq \frac{1 - \xi_S^{\max}}{\phi_I(1 - Q_{f,\text{AND}}(\tau))} \Rightarrow \Pr \left\{ 0.5(1 - \tau/T_S) \min \right. \\ &\quad \cdot \left. \{ R_{TR_k}(P_S h_{TR_k}), R_{R_kD}(P_{R_k} h_{R_kD}) \} > \text{Th}_S^{\text{req}} \right\} \\ &\geq \widehat{\omega}_{S,1}^{\text{CRT}} \Rightarrow \Pr \left\{ R_{TR_k}(P_S) > \frac{2\text{Th}_S^{\text{req}} T_S}{T_S - \tau} \right\} \Pr \\ &\quad \cdot \left\{ R_{R_kD}(P_{R_k}) > \frac{2\text{Th}_S^{\text{req}} T_S}{T_S - \tau} \right\} \geq \widehat{\omega}_{S,1}^{\text{CRT}} \\ &\Rightarrow \Pr \left\{ h_{TR_k} > \frac{(\widehat{\omega}_{S,2}^{\text{CRT}} - 1)\sigma_n^2}{P_S} \right\} \Pr \\ &\quad \cdot \left\{ h_{R_kD} > \frac{(\widehat{\omega}_{S,2}^{\text{CRT}} - 1)\sigma_n^2}{P_{R_k}} \right\} \geq \widehat{\omega}_{S,1}^{\text{CRT}} \\ &\Rightarrow \exp \left(\frac{(1 - \widehat{\omega}_{S,2}^{\text{CRT}})\sigma_n^2}{\lambda_{TR_k} P_S} + \frac{(1 - \widehat{\omega}_{S,2}^{\text{CRT}})\sigma_n^2}{\lambda_{R_kD} P_{R_k}} \right) \\ &\geq \widehat{\omega}_{S,1}^{\text{CRT}} \Rightarrow \ln \widehat{\omega}_{S,1}^{\text{CRT}} + (\widehat{\omega}_{S,2}^{\text{CRT}} - 1)\widehat{\omega}_{S,3}^{\text{CRT}} \leq 0. \end{aligned} \quad (\text{B.1})$$

So, the constraint (16).C2 is further equivalent to $\Omega(\tau) = \ln \widehat{\omega}_{S,1}^{\text{CRT}} + (\widehat{\omega}_{S,2}^{\text{CRT}} - 1)\widehat{\omega}_{S,3}^{\text{CRT}} \leq 0$, and the first derivation of $\Omega(\tau)$ can be formulated as

$$[\Omega]_\tau' = [Q_{f,\text{AND}}]_\tau' / (1 - Q_{f,\text{AND}}(\tau)) + \widehat{\omega}_{S,3}^{\text{CRT}} [\widehat{\omega}_{S,2}^{\text{CRT}}]_\tau', \quad (\text{B.2})$$

where $[\widehat{\omega}_{S,2}^{\text{CRT}}]_\tau' = 2 \ln(2) T_S R_S^{\text{req}} \widehat{\omega}_{S,2}^{\text{CRT}} W^{-1}(T_S - \tau)^{-2}$. The second derivative of $\Omega(\tau)$ is given by

$$[\Omega]_\tau'' = \widehat{\omega}_{S,3}^{\text{CRT}} [\widehat{\omega}_{S,2}^{\text{CRT}}]_\tau'' + \frac{[Q_{f,\text{AND}}]_\tau''}{1 - Q_{f,\text{AND}}(\tau)} + \left(\frac{[Q_{f,\text{AND}}]_\tau'}{1 - Q_{f,\text{AND}}(\tau)} \right)^2, \quad (\text{B.3})$$

where $[\widehat{\omega}_{S,2}^{\text{CRT}}]_\tau'' = 2(1 + \ln(2)) T_S \text{Th}_S^{\text{req}} W^{-1}(T_S - \tau)^{-1} [\widehat{\omega}_{S,2}^{\text{CRT}}]_\tau'$. We can easily examine $[\widehat{\omega}_{S,2}^{\text{CRT}}]_\tau'' > 0$ and $[\Omega]_\tau'' > 0$, which indicates $\Omega(\tau)$ is a concave function with respect to τ .

Data Availability

The data used to support the findings of this study are included within this article.

Conflicts of Interest

All the authors hereby declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was mainly supported by the National Nature Science Foundation of China (No. 61801319), Key Research Project of OUC (No. G18F3319Z), Educational Reform Project of Sichuan Provincial Department of Education (No. JG2018-1240), Education Reform Project of Sichuan Open University (No. XMKYC2020004Z), and Outstanding Young Teacher Program of OUC.

References

- [1] W. Yu, L. Musavian, and Q. Ni, "Tradeoff analysis and joint optimization of link-layer energy efficiency and effective capacity toward green communications," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3339–3355, 2016.
- [2] M. Liu, T. Song, J. Hu, J. Yang, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 641–653, 2019.
- [3] H. He and H. Jiang, "Deep learning based energy efficiency optimization for distributed cooperative spectrum sensing," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 32–39, 2019.
- [4] X. Liu, K. Zheng, L. Fu, X. Y. Liu, X. Wang, and G. Dai, "Energy efficiency of secure cognitive radio networks with cooperative spectrum sharing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 2, pp. 305–318, 2019.
- [5] L. Feng, Y. Kuang, B. Wu, Z. Dai, and Q. Yu, "Energy-efficient distributed spectrum sensing with combined censoring in cognitive radios," *IEICE Transactions on Communications*, vol. E99.B, no. 2, pp. 455–464, 2016.
- [6] J. Gui, Z. Li, and Z. Zeng, "Improving energy-efficiency for resource allocation by relay-aided in-band D2D communications in C-RAN-based systems," *IEEE Access*, vol. 7, pp. 8358–8375, 2019.
- [7] Z. Song, Q. Ni, K. Navaie, S. Hou, S. Wu, and X. Sun, "On the spectral-energy efficiency and rate fairness tradeoff in relay-aided cooperative OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6342–6355, 2016.
- [8] Y. Li, Z. Zhang, H. Wang, and Q. Yang, "SERS: social-aware energy-efficient relay selection in D2D communications," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5331–5345, 2018.
- [9] L. Feng, Y. Kuang, and B. Wu, "Energy-efficient configuration of power resource for OFDM-based opportunistic regenerative

- relay links,” *Chinese Journal of Electronics*, vol. 25, no. 2, pp. 341–347, 2016.
- [10] L. Feng, Y. Kuang, X. Fu, and Z. Dai, “Energy-efficient network cooperation joint resource configuration in multi-RAT heterogeneous cognitive radio networks,” *Electronics Letters*, vol. 52, no. 16, pp. 1414–1416, 2016.
- [11] G. Sharma and R. Sharma, “Energy efficient collaborative spectrum sensing with clustering of secondary users in cognitive radio networks,” *IET Communications*, vol. 13, no. 8, pp. 1101–1109, 2019.
- [12] L. Wang, M. Sheng, X. Wang, Y. Zhang, and X. Ma, “Mean energy efficiency maximization in cognitive radio channels with PU outage constraint,” *IEEE Communications Letters*, vol. 19, no. 2, pp. 287–290, 2015.
- [13] L. Zhang, M. Xiao, G. Wu, S. Li, and Y. C. Liang, “Energy-efficient cognitive transmission with imperfect spectrum sensing,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1320–1335, 2016.
- [14] X. Xu, J. Bao, H. Cao, Y. D. Yao, and S. Hu, “Energy-efficiency-based optimal relay selection scheme with a BER constraint in cooperative cognitive radio networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 191–203, 2016.
- [15] S. Maleki and G. Leus, “Censored truncated sequential spectrum sensing for cognitive radio networks,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 364–378, 2013.
- [16] Y. C. Liang, K. C. Chen, Y. Li, P. Mahonen, and D. Niyato, “Guest editorial - advances in cognitive radio networking and communications (I),” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 273–275, 2011.
- [17] Z. Dai, J. Liu, and K. Long, “Selective-reporting-based cooperative spectrum sensing strategies for cognitive radio networks,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3043–3055, 2015.
- [18] J. M. Moualeu, T. M. N. Ngatched, W. Hamouda, and F. Takawira, “Energy-efficient cooperative spectrum sensing and transmission in multi-channel cognitive radio networks,” in *2014 IEEE International Conference on Communications (ICC)*, pp. 4945–4950, Sydney, NSW, Australia, June 2014.
- [19] J. Wu, G. Wang, and Y. R. Zheng, “Energy efficiency and spectral efficiency tradeoff in type-I ARQ systems,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 356–366, 2014.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, 2013.

Research Article

Avoiding Spurious Retransmission over Flooding-Based Routing Protocol for Underwater Sensor Networks

Sungwon Lee , Yeongjoon Bae , Muhammad Toaha Raza Khan , Junho Seo ,
and Dongkyun Kim 

School of Computer Science & Engineering, Kyungpook National University, Daegu, Republic of Korea

Correspondence should be addressed to Dongkyun Kim; dongkyun@knu.ac.kr

Received 2 April 2020; Revised 30 June 2020; Accepted 1 August 2020; Published 25 August 2020

Academic Editor: Hideyuki Takahashi

Copyright © 2020 Sungwon Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In underwater wireless sensor networks (UWSN), acoustic communication naturally introduces challenges such as long propagation delay and high packet loss. The flooding-based routing protocol can address these challenges with its multipath characteristics. As in flooding-based routing, due to multipath propagation mechanism, not only DATA but also ACK messages are transmitted through multiple routes however still some packet loss will degrade the performance. So, to provide high reliability of message delivery, an efficient retransmission mechanism is inevitable. Though, if the network uses conventional transport layer protocol such as TCP, it will suffer a spurious retransmission problem as TCP was originally not designed for the multipath environment. In this paper, we propose route discrimination for flooding-based routing to reduce spurious retransmission in UWSN to solve the limitation. The notion of ACK copies waiting time (ACWT) is utilized which is selectively updated based on the similarity of paths of transmission of ACK message copies. We also improved our previous solution that lacks flexibility to cope with dynamic link error characteristics. Through evaluation, we verified that our new scheme achieves the performance improvements of 14%~84% in terms of retransmission ratio compared to the previous research.

1. Introduction

Nowadays, research interest for IoT (Internet for Things) technology has been increased to provide Internet technology to various “Things.” With the support of the IoT techniques, new IoT applications are developed that can make human lives more comfortable. The range of the “Things” includes vehicular, mobile and sensor devices. Underwater sensor nodes are one of the promising “Things” because of the various underwater applications. Those applications include seawater temperature monitoring, underwater pollution monitoring, and coastal threat monitoring, developed to serve the ecosystem and human lives [1–6].

UWSNs have different characteristics from the terrestrial wireless sensor networks as they commonly use acoustic waves instead of radio waves. Due to the characteristics of the acoustic waves, underwater communication suffers from various limitations including slow propagation, high error rate [7–11]. To overcome these limitations and to achieve

high throughput, various protocols are newly proposed for UWSNs. Among them, flooding-based routing protocols are proposed to transmit data packets from source to the destination reliably. Unlike conventional routing protocols designed for the terrestrial networks such as ad hoc on-demand distance vector (AODV) [12], the flooding-based routing protocols have two advantages for error-prone and slow UWSNs [13]. Since the flooding-based routing protocols send multiple copies of a data packet along different paths, the original packet could be received at the sink even if few copies of the packets are lost. Moreover, the source node could send a data packet without the path establishment in advance. Detailed descriptions for the flooding-based routing protocols are introduced in Section 2.1.

Although, these routing protocols can improve the reliability but still lack far behind from meeting the IoUT (Internet of Underwater Things) application requirements. For example, in some IoUT applications, large and sensitive data (oil pipeline leakage image) is to be transmitted between the

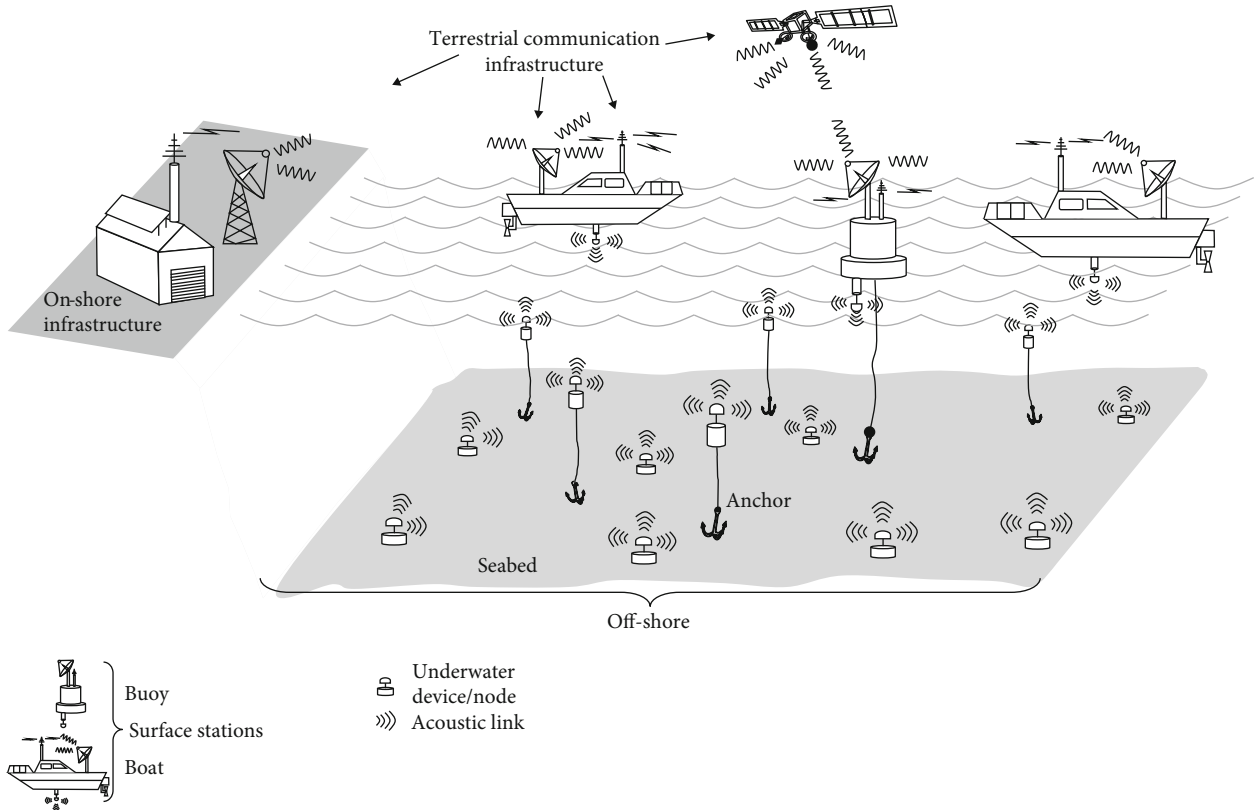


FIGURE 1: UWSN architecture.

sensor node and the sink node (Figure 1). Moreover, similar to software-defined networking architectures, if files which are used to upgrade the sensor node firmware should be transmitted, all sensor nodes have to receive these files and as even one bit of error is not allowed [14]. To support this packet transmission, reliable transport protocols which can detect the packet loss and support packet retransmission operation are required over the flooding-based routing protocols.

One of the most famous transport protocols which provides packet loss detection and retransmission operation is the transmission control protocol (TCP) [15]. In the TCP, the source set retransmission timeout (RTO) based on round trip time (RTT). If the source node cannot receive the acknowledgment (ACK) message from the destination node, it retransmits the packet [16]. Hence, the RTO calculation is one of the most important operations in TCP. If the calculated RTO is too long, throughput could be decreased because long time is required for retransmissions. On the other hand, if the RTO expires before ACK is received, a spurious retransmission problem could occur. The spurious retransmission wastes the bandwidth since the source node retransmits data packet which is successfully transmitted at the destination node. In the terrestrial networks, the source uses various RTO calculation algorithms which are well-operated over single-path routing protocols.

However, as mentioned above, IoUT devices use the flooding-based routing protocols which transmit packets (both of data and ACK) along multipaths. Over these routing

protocols, even though an ACK message is transmitted at the destination, the source node will receive multiple ACK messages which are flooded along different paths. If the traditional RTO calculation algorithms such as Khan's algorithm are implemented, the source updates RTT when the first ACK is received. Other ACKs will be ignored because the source node already knows that the destination node received the corresponding data packet. Therefore, the source node will set RTO to the time required for the first received ACK to arrive. When the source node uses this RTO for retransmission and the first ACK is lost during transmission, the source will perform packet retransmission (although another ACK messages which are transmitted along different paths) as described in Figure 2. Therefore, in order to avoid the spurious retransmission problem over the flooding-based routing protocol, the new RTO calculation which reflects another ACK messages which are received after the first ACK reception method is required.

In other words, RTO should be calculated in a way such that, for each data transmission, at least one of the many ACK packets transmitted over multiple paths can be received successfully. In our previous work [17], we proposed a technique that uses the similarity of paths from which packets have been sent to identify ACK packets that have a low probability of loss. The scheme also updates the RTO selectively where path similarity is defined in relationship with the same nodes that both paths share. However, in our previous work, a number of shared ancestor nodes among multiple paths are used only to determine whether to reflect the RTT of the

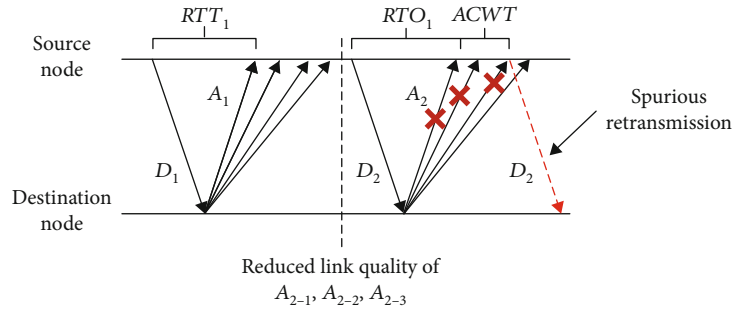


FIGURE 2: Spurious retransmission.

newly received ACK message or not. In other words, if the number of nodes shared by both paths is less than the threshold, the RTTs of all ACK messages are equally reflected. Due to this feature, a path which has poor link quality, or a large number of nodes could change the RTO significantly, thereby reducing throughput.

Hence, in this paper, we propose a technique to calculate the RTO that adapts to the current network situation by considering the information of each path and by applying path similarity analysis. We simulated and compare the proposed scheme using a C program with conventional RTO calculation mechanism [18] and previous work [17].

The rest of this paper is organized as follows. In Section 2, related works are discussed. Section 3 contains the detail description of the proposed techniques to solve the problems mentioned in this paper, while Section 4 analyses the performance of the proposed scheme through simulation comparison. Finally, Section 5 concludes our work.

2. Related Work

2.1. Flooding-Based Routing Protocols. As mentioned above, in flooding-based routing protocols, nodes forward every received packet without defining a separate path [19–21]. In these routing protocols, each node establishes a flooding area that directs its location information to the sink node. Then, whenever each node transmits a data packet, information related to the flooding area is recorded in the header of the data packet and broadcasted. If the neighbouring nodes of the transmitting nodes receive this packet, it confirms that they are included in the flooding area. If it is included in the flooding area, it updates the header information in the same way and broadcasts the packet. If it is not included in the flooding area, discard the packet.

Among various flooding-based routing protocols, the depth-based routing (DBR) protocol [20] is a representative flooding-based routing protocol. In the DBR, when the sender node transmits a packet, its depth information is included in the routing header. On arrival of a packet, nodes located shallower than the sender node operated as relays and retransmit it towards the surface. Hence, the DBR not only allows packets to be delivered away from the sink but also has a problem that the size of the flooding area is always fixed regardless of link quality.

To overcome this problem, in the Directional Flooding-based Routing (DFR) [21] protocol, the size of the flooding area is adjusted by the average link quality. The transmitting node measures the average ETX (Expected Transmission Count) with neighbouring nodes. Each time a packet is transmitted, the sender node records its location information and the base angle, which increases as the average ETX value becomes lower, in a data packet and broadcasts it. The receiving node computes the angle between the vector connecting itself and the sender node and the vector connecting itself and the sink node. If the calculated angle is less than the base angle included in the packet, the node transmits the packet. Therefore, the worse the link quality, the more nodes participate in the packet transmission and improve the reliability. Conversely, the better the link quality, the fewer nodes can participate in packet transmission and save network resources [22].

2.2. Retransmission Technique. The transport layer protocols provide an end-to-end connection between a sender and a receiver. These protocols are classified into two types: user datagram protocol (UDP) [23] and transmission control protocol (TCP) [15]. UDP has a very simple transmission method without any guarantee of successful delivery. In TCP, it confirms sender nodes regarding successful packet delivery. Since in UDP, the transmission overhead is small as compared to TCP, it is preferable in real-time applications like voice over internet protocol (VoIP) and real-time streaming. TCP, on the other hand, establishes a connection between a sender and a receiver before data transmission and transmits data reliably by providing functions such as acknowledgment, retransmission, and message reordering in data transmission. Therefore, it is mainly used in the web, e-mail transmission, or file transfer application where data must be normally transmitted. The TCP has multiple versions such as TCP-Tahoe, TCP-Reno, TCP-SACK [24], TCP-New Reno [25], TCP-Vegas [26], and TCP-Westwood [27] are available. Each version targets different performance matrix for operation such as congestion control, packet transmission rate control, packet loss, and duplication problems while operating following the core TCP.

CoAP (Constrained Application Protocol) is a lightweight transport protocol for constrained nodes (CN). Taking into account the limited computing power of CNs, CoAP is designed to work over UDP instead of TCP.

However, since UDP is a protocol that cannot provide end-to-end reliability, CoAP can use the ACK message in the form of a UDP packet. A packet requesting transmission of an ACK packet is defined as a Confirmable message, and a packet not requesting transmission of an ACK packet is defined as a nonconfirmable message (the type of this message is determined by the upper layer such as application). If the source node sends a confirmable message, it sets a retransmission timer (RTO) likely TCP. If the corresponding ACK message is not received before the expiration of the retransmission timer (due to delay or loss), the source node retransmits the confirmable message. Hence, appropriate RTO calculation is needed for not only TCP but also CoAP.

3. Proposed Scheme

In this section, after defining the problem, the proposed scheme is introduced. In the proposed scheme, we assume that DFR and Khan's algorithm in TCP-Reno are used for flooding-based routing protocol and retransmission algorithm, respectively. This is because DFR is a routing protocol that not only improves reliability but also saves network resources, and Khan's algorithm is a proven retransmission algorithm in various networks such as the Internet.

3.1. Problem Definition. In our previous work [17], ACWT is selectively updated considering the path similarity between received ACK copies. ACWT is calculated systematically to incorporate randomly arranged network topology while reducing unnecessary retransmission.

However, in our previous work, the maximum number of ACK copies (ACK_{max}) is set as a fixed value. We assumed that the quality of packet transmission is constant for all links, and only the difference in arrival time of ACK copies is considered. However, in real underwater networks, link qualities of each path changes dynamically. The number of ACK messages received for each segment of one flow changes. ACWT is calculated based on the received amount of the previous ACK message makes it impossible to accurately set the retransmission timer of this ACK message. So, we utilize the ratio of the successfully delivered packet over the total number of packets transmitted between two nodes as a parameter to depict the link quality between the two nodes.

If the ACWT is calculated using the fixed ACK_{max} without considering the transmission success rate of the ACK copies transmitted to the source node, it cannot cope with the dynamic changes in the underwater wireless communication channel. One of the scenarios is depicted in Figure 2 in which a sudden decrease in the link quality of the path leads to spurious retransmission. For ACWT calculation, ACK_{max} is continuously updated in the new RTO calculation technique based on the instantaneous network conditions.

In this paper, we propose a method in which the source nodes manage the information of a distinguished path by using bitmap operation while keeping a separate table. In this way, ACK copies transmitted to the source node form different paths. For every individual path based on the average RTT calculation, the change in the packet transmission suc-

cess rate is monitored. As a result, the source node continuously manages the average RTT and the average packet transmission success rate of different paths formed between itself and the destination node. The source node adjusts the ACK_{max} flexibly based on the average RTT and the average packet transmission success rate.

3.2. Proposed Solution. In the common IoUT, if nodes have low remaining energy or have a high computing demand sleep to save energy. Otherwise, if nodes want to communicate with each other, they are awake and start idle listening to receive to transmit or relay data packets. The sender node that intends to send some collected data is termed as the source node. Every sensor node creates and stores a separate path information table for the path information distinguished through the bit map operation and continuously manages this information for the packet transmission. Every node manages information regarding travelled section (TS) field representing the corresponding path followed by the packet, an average RTT of ACK copies received through the corresponding path and an average transmission success rate of ACK copies. By managing this information, the source node can know how many routes are formed between the node itself and the destination node, and how much the expected success rate of ACK copy is obtained when the value of ACK_{max} is changed. Algorithm 1 illustrates the pseudocode of the path distribution-based information update along with ACWT calculation.

First, when a new ACK copy (ACK_{recv}) is received, the similarity between the TS fields stored in the current path table and the TS field (TS_{recv}) of the corresponding ACK copy is calculated to determine whether a similar path is already managed. If it is determined that the TS_{recv} is similar to the TS field stored in the route table, it is determined that the ACK copy is transmitted through the same route, and the average RTT of the route is updated by the RTT obtained through ACK_{recv} . On the other hand, if TS_{recv} is not similar to any TS field stored in the current path table, it is determined that the ACK copy is transmitted through the new path. The Num_{path} is incremented by 1, and TS_{recv} is stored as the TS field of the new path in the path table. Then, the source node calculates the expected packet transmission success rate of each path based on the number of times that the source node has transmitted the data packet and the number of ACK copies received through each path. When the calculation of the average RTT and the average expected packet transmission success rate of all paths is completed through this operation, the ACK_{max} is calculated.

3.3. Calculation of the ACK_{max} . In algorithm 1, the utility function $U(i)$ is one of the most important factors to determine a new round trip time measured by the i th ACK (namely $RTT[i]$) is an appropriate value to update of ACWT. To calculate the $U(i)$, the equations (1) and (2) are used as follows.

$$V(i) = \omega time \times \left(2 - \frac{RTT[i] - RTT[i-1]}{RTT[0]} \right). \quad (1)$$

```

1:receive new ACK(ACKrecv)
2:decide ← true
3:for i=0; i < Numpath; i++ do
4:  if PS(TSpath[i], TSrecv) > PSthresh then
5:    decide ← false, update RTTpath[i]
6:  end if
7:end for
8:if (decide) then
9:  Numpath ← Numpath + 1, TSpath [Numpath] ← TSrecv
10:end if
11:for i=0; i < Numpath; i++ do
12:  update RTTpath[i]
13:end for
14:for i=0; i < Numpath; i++ do
15:  if U(i) > Uthresh then
16:    ACK ← i + 1
17:  end if
18:end for
19: for i=0; i < ACKmax; i++ do
20:  calculate Tack and update ACWT with RTTpath[i]
21:end for

```

ALGORITHM 1. Path distinction-based information update and ACWT calculation.

In this equation (1), $V(i)$ and ω time are a vector value of $RTT[i]$ and a weight factor for $V(i)$, respectively. After calculating the $V(i)$, the $U(i)$ is calculated as follows.

$$U(i) = V(i) + \omega \text{prob} \times (1 - (1 - PDR[i]) \cdots (1 - PDR[0])) - PDR[0]. \quad (2)$$

In this equation (2), $PDR[i]$ and ω prob are an expected packet delivery ratio of i th path and a weight factor for $U(i)$, respectively, when considering the i th path, compared with the case considered only up to the $(i-1)$ th path. The larger the difference between the i th ACK arrival path and the $(i-1)$ th arriving RTT path, the smaller the value of the term itself. Instead of using the absolute difference between these RTTs, that difference is normalized by the $RTT_{path}[0]$ to be proportional to the RTT of the first arriving ACK. Therefore, the first term has a larger value as the difference between the i th ACK arrival path and the $(i-1)$ th arrival path RTT becomes smaller. The second term is used for calculating the difference in the expected PDR when considering the i th path, compared with the case considered only up to the $(i-1)$ th path. The larger the difference between the expected packet transmission success rate when considering the i th path and the expected packet transmission success rate when considering up to the $(i-1)$ path, the greater the value of the term itself.

Therefore, the equation (2) calculates the appropriate value of $U(i)$ how the i th path to be reflected in the ACWT update through the time difference that occurs when considering the i th path and the increase of the packet transmission success rate. In addition, the equation compares the value

with a threshold value (U_{thresh}) and finally obtains the optimal performance in the current path situation.

In this paper, we set these 3 network parameters (ω time, ω prob, and U_{thresh}) to 0.25, 0.5, and 0.45, respectively. These values are measured by repeated simulations.

3.4. Demonstration. The functioning of the proposed technique will be exemplified by a simple case depicted in Figure 3. As shown in the figure, four paths exist between the source node and the sink node.

In this case, the average RTT and link quality for each path in (a) and (b) are changed as shown in Table 1.

First, RTT increases gradually in the order of Path0, Path1, Path2, and Path3 in (a) and (b). In (a) situation, the link quality of Path0, Path1, and Path2 was 85%, and Path3 alone had a relatively low link quality of 40%. In each situation, $U(i)$ is calculated as shown in Table 2.

Therefore, in (a), $U(1)$ is calculated to be higher than U_{thresh} and the optimal ACK_{max} is calculated as 2. However, in (b), when the RTT of Path0 increases by 0.1 seconds and the transmission quality drops to 40%, the optimal ACK_{max} is calculated to be 3 because $U(2)$ is larger than U_{thresh} . Therefore, if the transmission quality of Path0 drops, it can be confirmed that in addition to Path1, ACK can be further awaited by reflecting the addition of Path2, which was not previously reflected, to the calculation of ACWT.

4. Performance Evaluation

4.1. Simulation Environments. To evaluate the performance of our scheme, we used NS-2.34 simulator and underwater acoustic signals propagation model. Other simulation parameters were set as shown in Table 3. Moreover, to evaluate the contribution of the proposed scheme only, we

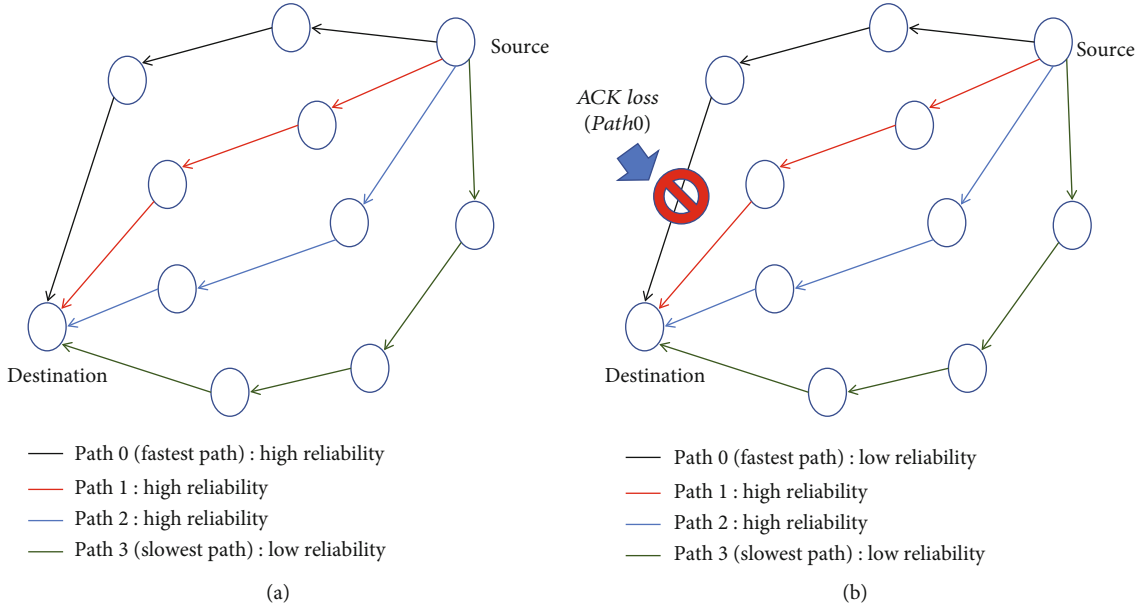


FIGURE 3: Example Scenario.

TABLE 1: RTT and link quality.

Path	RTT	(a)	RTT	(b)
		Link quality		Link quality
Path0	5.0 sec	85%	5.1 sec	40%
Path1	5.3 sec	85%	5.3 sec	85%
Path2	6.5 sec	85%	6.5 sec	85%
Path3	7.6 sec	40%	7.6 sec	40%

TABLE 2: $U(i)$ values.

$U(i)$	(a)	(b)
$U(1)$	0.548	0.745
$U(2)$	0.449	0.479
$U(3)$	0.445	0.448

TABLE 3: Simulation parameters.

Parameter	Value
Transmission range	250 meters
Routing protocol	DBR [14]
ACK_{max}	2
PS_{thresh}	0.25
ω_{time}	0.25
ω_{prob}	0.5
U_{thresh}	0.45

selected the original TCP instead of other TCP versions. If the proposed technique is applied to other TCPs, it is not possible to verify whether the improved performance is due to the proposed technique or another technique.

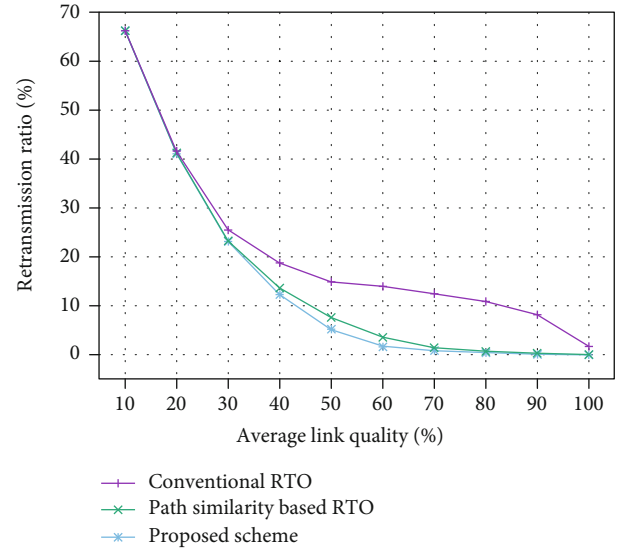


FIGURE 4: Retransmission ratio in uniform LQ scenario.

The size of the network is set to 1600×800 m. The whole network is divided into 8 rows and 4 columns resulting in a total of 32 sections. The node nearest to the point $(0, 800)$ is designated as the source node, and the node closest to the point $(1600, 0)$ is designated as the sink node. The nodes were relocated randomly in each experiment run. In addition, DBR is used as the routing protocol.

4.2. Performance Factor. We assume two different situations for performance evaluation. First, the link quality (LQ) of the entire network is uniform. This is to evaluate the performance of the proposed technique for the comparison with the existing method (path similarity-based RTO). In second, the link quality of the entire network

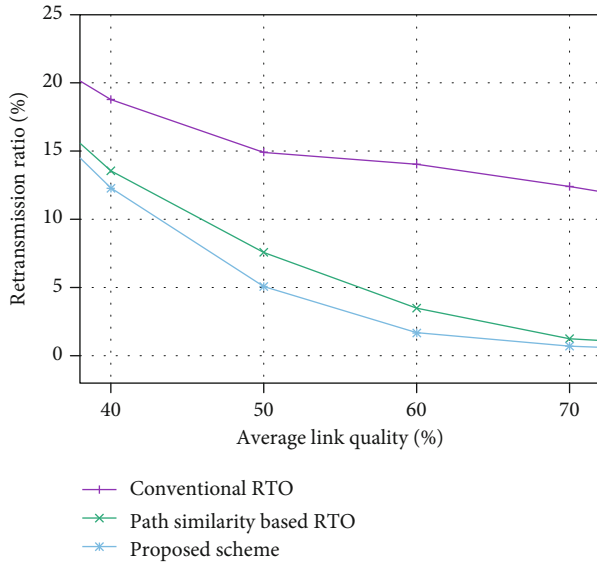


FIGURE 5: Retransmission ratio in uniform LQ scenario (40~70 nodes).

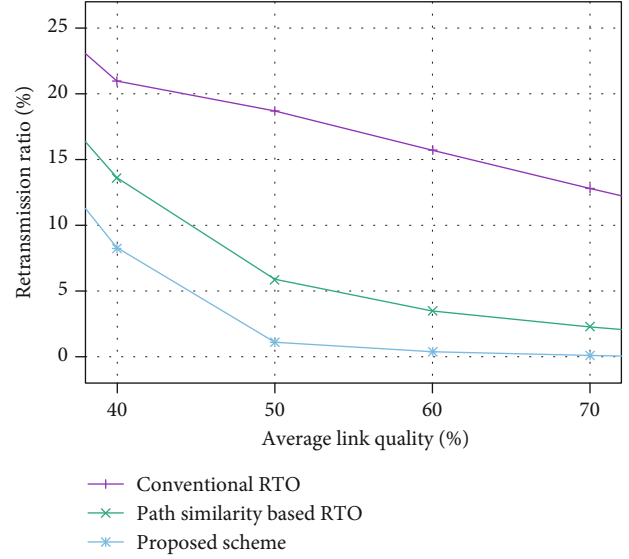


FIGURE 7: Retransmission ratio in different LQ scenario (40~70 nodes).

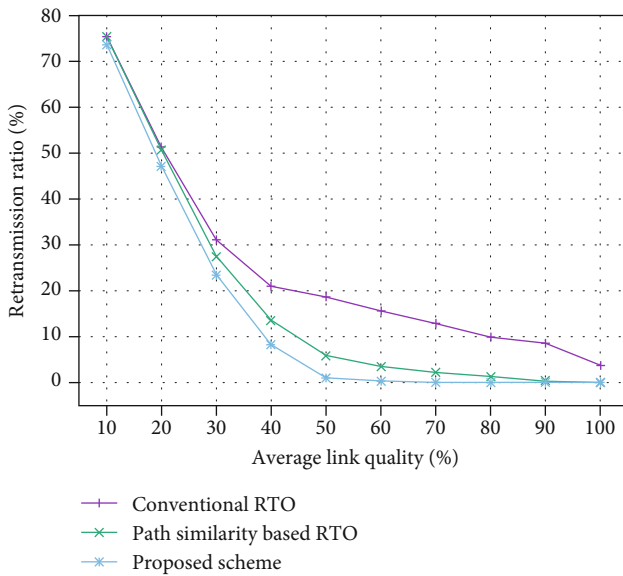


FIGURE 6: Retransmission ratio in different LQ scenario.

is nonuniform that varies from time to time. We can evaluate the performance of the proposed method by comparing it with that of the existing method. The evaluation parameters include “retransmission ratio” that is the number of packets retransmitted by the source node divided by the total number of packets transmitted by the source node, where “average link quality” indicates the average transmission success rate of the entire network.

4.3. Performance Comparison in Uniform Link Quality of Network. In this experiment, we first set the average link quality of all the links uniformly, and experimentally examined how the retransmission ratio changes according to the change of the average link quality. As shown in Figure 4,

the retransmission ratio increases as the average link quality decreases regardless of the RTO calculation technique used. Especially, it can be confirmed that the proposed algorithm based on the path similarity-based RTO that can additionally wait for the ACK copy and the proposed scheme achieves a lower retransmission ratio than the conventional RTO calculation method regardless of the change in link quality.

The performance difference between the two schemes with ACWT and the existing RTO calculation technique starts from when the average link quality is 100% and gradually increases from 70% until the average link quality becomes lower than 60%. All converge to one point. This is because, when the average link quality is high, it is possible to reduce the retransmission by using additional received ACK copies. As the link quality decreases, ACK copies are also lost as a result successful ACK copy cannot arrive.

However, as shown in Figure 5, the difference in performance between the path similarity-based RTO calculation technique and the proposed technique is conspicuous in the interval of 70%~40% of the average link quality. This is because the fixed ACK_{max} , the proposed method in this paper seems to be due to the fact that the expected rate of packet transmission is low for all the paths at present, so that the ACK_{max} is made large and waiting for the maximum number of ACK copies.

4.4. Performance Comparison with Different Link Qualities.

In this experiment, the link quality of the entire network is not uniform, and the link quality of each link is constantly changed. As can be seen in Figures 6 and 7, the retransmission ratios are increased as the average link quality decreases regardless of the RTO calculation technique used in this experiment.

However, unlike the experimental results in Figure 6, it can be seen that the performance difference between the proposed method and the path similarity-based RTO calculation

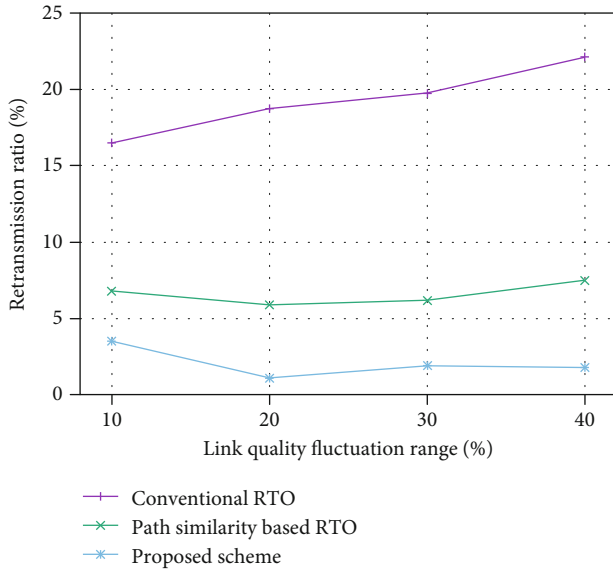


FIGURE 8: Retransmission ratio in LQ fluctuation scenario.

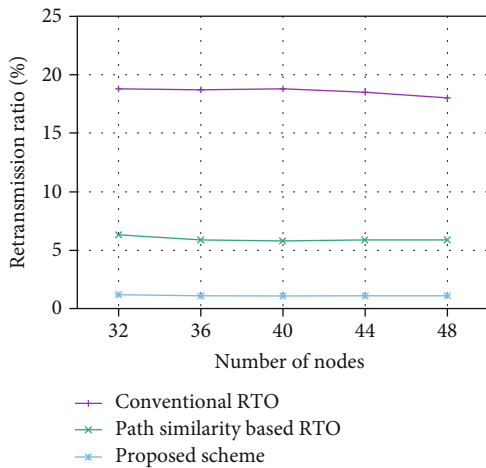


FIGURE 9: Retransmission ratio in LQ fluctuation scenario (32~48 nodes).

technique is noticeable regardless of the average link quality. It monitors the state of multipath formed between the source node and the destination node, measures the RTT and the expected transmission success rate for each path, and calculates the ACK_{max} value that can achieve the optimum performance based on this. This is a result of proving that the operation is effective in the same environment as the experiment in which the link quality fluctuates. It is shown that the proposed method improves performance by at least 14% and up to 94% compared with the existing similarity-based RTO calculation method.

Also, in this experiment, the performance difference is conspicuous in the case where the average link quality is higher than a certain level, but the performance difference is decreased as the average link quality is lowered. This is because, as mentioned in the previous experimental results,

when the average link quality is high, it is possible to reduce the retransmission by using additional received ACK copies. However, as the link quality is lowered, ACK copies are also lost and waiting for ACK copies. This is because it is not possible to successfully receive an ACK copy.

After fixing the average link quality at 50%, we experimented the change of retransmission ratio according to the link quality fluctuation width by changing the fluctuation range of the link quality of the whole network. As shown in Figure 8, the fluctuation width is changed from 10% to 40%, and the retransmission ratio is generally increased as the variation of the link quality is larger regardless of the RTO calculation method used. Regardless of the fluctuation of the link quality, the proposed scheme achieves a lower retransmission ratio than the other two schemes.

Also, the retransmission ratio increases linearly with the increase of the link quality fluctuation in the case of the conventional RTO calculation technique. However, the path similarity-based RTO technique and the proposed technique do not. This seems to be caused by the ACWT, which is commonly used by both schemes. It is because the link quality of the multipath is changed differently, and k and TACK are not calculated stably.

Finally, after changing the average link quality to 50% and the link quality fluctuation to 20%, we experimented with changing the number of nodes in the network and changing the retransmission ratio. As can be seen in Figure 9, the number of nodes varied from 32 to 48. The results show that the proposed method shows the best performance irrespective of the number of nodes. In particular, in the case of the conventional RTO calculation technique, a slight improvement in performance was confirmed as the number of nodes increased. However, in the case of the other two techniques, no noticeable performance change was observed even when the number of nodes increased. This is because the ACWT is calculated through the routes with a similar degree of similarity even if more multipath are formed as the number of nodes increases.

5. Conclusion

In this paper, the information of the multipaths established between the source and the destination node is utilized over traditional TCP to avoid spurious retransmission in UWSN while using flooding-based routing. In the proposed scheme, the source node retransmission timeout (RTO) calculation measures an expected number of ACK messages (ACK_{max}) based on the reliabilities of paths. Using ACK_{max} , the proposed scheme reduces the unnecessary retransmission. While waiting for the ACK of the maximum number of paths, ACK_{max} is increased, as the success rate of ACK transmission of each path is lower. If ACK-transmission rate is high, ACK_{max} is decreased enabling faster retransmission. Through performance evaluation via multiple simulations, the proposed scheme achieves 14%~84% improved performance in the form of retransmission ratio.

In future work, we will implement our proposed scheme to real underwater communication devices such as Benthos modem. Based on the real test result, we will

improve our proposed scheme. Moreover, we will work on an underwater congestion control method utilizing the concept of consistency.

Data Availability

All data used to support the findings of the study is included within the article.

Conflicts of Interest

None of the authors have any conflicts of interest.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2016R1D1A3B01015510).

References

- [1] S. Mustafa, M. A. Khan, and S. A. Malik, "Design of underwater sensor networks for water quality monitoring," *World Applied Sciences Journal*, vol. 17, no. 11, 2012.
- [2] F. Raimondi, M. Trapanese, P. Martorana, V. Franzitta, and V. A. Di Dio, "Security and surveillance application of the innovative monitoring underwater buoy systems (MUnBuS) on the protected marine area (AMP) of capo gallo (PA-IT)," in *OCEANS'15 MTS/IEEE*, Washington, 2015.
- [3] K. Lam, R. S. Bradbeer, P. K. Shin, K. K. Ku, and P. Hodgson, "Application of a real-time underwater surveillance camera in monitoring of fish assemblages on a shallow coral communities in a marine park," in *IEEE OCEANS*, Vancouver, Canada, 2007.
- [4] C. C. Kao, Y. S. Lin, G. D. Wu, and C. J. Huang, "A comprehensive study on the internet of underwater things: applications, challenges, and channel models," *Sensors*, vol. 17, no. 7, article 1477, 2017.
- [5] G. Xing, Y. Chen, L. He et al., "Energy consumption in relay underwater acoustic sensor networks for NDN," *IEEE Access*, vol. 7, pp. 42694–42702, 2019.
- [6] E. Liou, C. Kao, C. Chang, Y. Lin, and C. Huang, "Internet of underwater things: challenges and routing protocols," in *Proc. of 2018 IEEE International Conference on Applied System Invention (ICASI)*, pp. 1171–1174, Chiba, April 2018.
- [7] M.-S. Kim, S.-W. Lee, M. T. R. Khan et al., "A new CoAP congestion control scheme using message loss feedback for IoUT," in *Proc. of the 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19)*, pp. 2385–2390, Limassol, Cyprus, April 2019.
- [8] J. Seo, S. Lee, M. T. R. Khan, and D. Kim, "A new CoAP congestion control scheme considering strong and weak RTT for IoUT," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing (SAC '20)*. Association for Computing Machinery, pp. 2158–2162, New York, NY, USA, 2020.
- [9] G. Han, X. Long, C. Zhu, M. Guizani, and W. Zhang, "A high-availability data collection scheme based on multi-AUVs for underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1010–1022, 2020.
- [10] H. U. Yildiz, "Maximization of underwater sensor networks lifetime via fountain codes," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4602–4613, 2019.
- [11] I. F. Akyildiz, P. Wang, and Z. Sun, "Realizing underwater communication through magnetic induction," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 42–48, 2015.
- [12] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," Technical report, 2003.
- [13] S. LEE and D. KIM, "aCK loss-aware RTO calculation algorithm over flooding-based routing protocols for UWSNs," *IEICE Transactions on Information and Systems*, vol. E97.d, no. 11, pp. 2967–2970, 2014.
- [14] J. Postel and J. Reynolds, "File transfer protocol," Technical report, 1985.
- [15] J. Postel, "Transmission control protocol," Technical report, 1981.
- [16] M. A. J. C. Paxson Vern and M. Sargent, "Computing TCP's retransmission timer," Technical report, 2011.
- [17] Y. Bae, S. Lee, Y. Jeong, and D. Kim, "Path similarity based spurious retransmission minimization over flooding based routing in UWSN," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, Prague, Czech Republic, July 2018.
- [18] V. Jacobson, "Congestion avoidance and control," *ACM SIGCOMM Computer Communication Review*, vol. 18, no. 4, pp. 314–329, 1988.
- [19] J. H. C. Xie Peng and L. Lao, *VBF: Vector-Based Forwarding Protocol for Underwater Sensor Networks*, International conference on research in networking, 2006.
- [20] Z. J. S. Yan Hai and J. H. Cui, *DBR: Depth-Based Routing for Underwater Sensor Networks*, International conference on research in networking, 2008.
- [21] D. Shin, D. Hwang, and D. Kim, "DFR: an efficient directional flooding-based routing protocol in underwater sensor networks," *Wireless Communications and Mobile Computing*, vol. 12, no. 17, 1527 pages, 2012.
- [22] D. S. J. De Couto, "High-throughput routing for multi-hop wireless networks," PhD thesis at Massachusetts Institute of Technology, 2004.
- [23] J. Postel, "User datagram protocol," Technical report, 1980.
- [24] K. Fall and S. Floyd, "Simulation-based comparisons of Tahoe, Reno and SACK TCP," *ACM SIGCOMM Computer Communication Review*, vol. 26, no. 3, pp. 5–21, 1996.
- [25] T. H. Floyd Sally and A. Gurtov, *The NewReno Modification to TCP's Fast Recovery Algorithm*, Technical report, 2004.
- [26] L. S. Brakmo and L. L. Peterson, "TCP Vegas: end to end congestion avoidance on a global internet," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 8, pp. 1465–1480, 1995.
- [27] S. Mascolo and R. Wang, "TCP westwood: bandwidth estimation for enhanced transport over wireless links," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, New York, USA, 2001.

Research Article

A Simple and Energy-Efficient Flooding Scheme for Wireless Routing

Satoshi Yamazaki ¹, Yu Abiko,² and Hideki Mizuno³

¹Department of Control and Computer Engineering, National Institute of Technology, Numazu College, Numazu 410-8501, Japan

²Graduate School of Integrated Science and Technology, Shizuoka University, Hamamatsu 432-8011, Japan

³School of Engineering, Tokai University, Isehara 259-1193, Japan

Correspondence should be addressed to Satoshi Yamazaki; s-yamazaki@numazu-ct.ac.jp

Received 3 April 2020; Revised 3 July 2020; Accepted 15 July 2020; Published 20 August 2020

Academic Editor: Yujin Lim

Copyright © 2020 Satoshi Yamazaki et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In many wireless networks, such as ad hoc, sensor, and delay-tolerant networks, the destination node is determined from the source node by the flooding process. Flooding efficiency is important for nodes, because they are driven by limited batteries. In this paper, we propose a simple flooding scheme to transmit a route request (RREQ) message based on the remaining power of its own node without using control packets and complex calculations. We applied the proposed scheme to ad hoc on-demand distance vector (AODV) routing protocol as an example and carried out computer simulations (ns3). The results showed that the proposed scheme was superior to conventional schemes in static and mobile scenarios. First, we showed the limit of node density that causes the decrease of throughput in the proposed scheme and that the proposed scheme was superior in terms of energy efficiency (bits/J), including throughput and energy consumption. Next, as the number of flooding times is made uniform in the proposed scheme, all nodes will have almost the same battery replacement time. As a result, when the nodes are static, the lifetime in the proposed scheme is longer than that in the conventional scheme.

1. Introduction

The internet of things (IoT) allows users to select the type of wireless communication methods, such as low-power wide-area (LPWA), Wi-Fi, and short-range communications based on user requirements and system specifications. Even when a base station is not operational, nodes can communicate with each other without the base station [1]. To enable communication even in such a situation, a mobile ad hoc network (MANET) based on multihop communication between many nodes is used without relying on centralized infrastructure [2–4]. In this paper, we focus on MANET. One of the critical challenges to address for MANET and sensor networks is to minimize their power consumption since they are configured with battery-powered nodes [5]. This is because in a MANET consisting of many nodes, information cannot reach the destination node if a node runs out of battery power while transmitting and receiving information. Though the flooding process is necessary for route establishment, the

battery life is decreased if the flooding process is focused on specific nodes. In such a case, it is difficult to replace the battery. Therefore, reducing the energy consumption from the physical layer to the network layer is necessary from the viewpoint of both software and hardware. Several surveys on the energy efficiency of MANET, mainly with regard to the routing protocols, have been described [6–10]. In addition, various protocols for sensor networks based on MANET have been developed [11–13].

In this paper, we propose a simple flooding scheme using only the remaining energy of each node, without the need for additional information. In general, there is a trade-off between energy consumption and throughput improvement [14], and even though energy consumption can be improved, the throughput may decrease. In previous schemes, energy and transmission performance, such as throughput and delay, are evaluated separately, and the evaluation has not been always adequate. Therefore, we evaluate the proposed scheme using the energy efficiency (bits/J) [15], which is an

important index in next-generation wireless communication, combining with energy consumption and throughput. In addition, we show the dominance of the proposed scheme by comprehensively evaluating the variance of the remaining energy and lifetime of the whole network when a node is static or moving. Though the proposed scheme is widely applicable to flooding-based wireless routing in optimized link state routing (OLSR) [16] in MANET, wireless sensor network (WSN), delay-tolerant networking (DTN) [17], and so on, in this paper, ad hoc on-demand distance vector (AODV) [18] routing in MANET is used as an example.

The main contributions of this paper are as follows. In this paper, we proposed a simple flooding scheme to transmit RREQ messages based on only the remaining power of its own node in the flooding process without the other information such as control packets and complex calculations. The proposed scheme was applied to AODV of MANET routing, and we showed that the proposed scheme was superior to conventional methods [18, 26] in the following points using computer simulations:

- (i) As the node density decreases, the throughput of the proposed scheme is lower compared with that of the conventional methods, whereas the throughput of the proposed scheme increases when the node density becomes higher. The limit of node density that causes the decrease of throughput is nearly 40 nodes/km² in this evaluation
- (ii) Irrespective of the node density, the energy consumption was reduced in the proposed scheme
- (iii) As a result, the proposed scheme is shown to be superior in terms of energy efficiency (bits/J), with regard to throughput and energy consumption
- (iv) As the nodes to be flooded are equalized in the proposed scheme, all nodes have almost the same battery replacement time
- (v) When the nodes are static, the network lifetime in the proposed scheme does not decrease regardless of the node density or the number of flows

The rest of this paper is organized as follows. In Section 2, we describe the related works. In Section 3, we show the proposed scheme, smart AODV, which is called SAODV. In Section 4, we discuss the performance of the proposed scheme using computer simulations. Finally, we present our conclusions in Section 5.

2. Related Works

Energy consumption reduction methods in MANET are mainly divided into two groups: energy-efficient route selection and flooding efficiency. The latter related to the proposed scheme is described in Section 2. For the former, methods have been proposed in which the amount of remaining energy of a node is used as a route selection metric [19–21]. Though the lifetime of the entire network can be extended, the regular exchange of control messages increases

the additional energy consumption. It is a serious problem in an environment when the network topology changes every moment. If the battery level of a node falls below a certain threshold, a method of exiting the route [22] or avoiding participation in route construction [23] has been proposed. It is difficult to determine an appropriate threshold, and new items are required for the control packets, making the processing more complicated.

Next, we describe conventional researches on energy consumption based on efficient flooding. First, the flooding operation is briefly described using AODV routing as an example [21]. In AODV, the source node *S* transmits a route request (RREQ) packet to neighboring nodes to determine the route. The node that received the RREQ packet repeats the transmission of the RREQ packet in the same manner so that the RREQ packet finally reaches the destination node *D*. In addition to this route discovery operation, flooding is also performed in the operation of searching for the cost route of the disconnected route or route maintenance. As well as AODV, the same process can be applied to the other on-demand protocol such as dynamic source routing (DSR) [24]. In OLSR [16], which is a proactive protocol, multipoint relaying (MPR) nodes serve as the flooding nodes. The problem with flooding is that a sudden increase in traffic causes many packet collisions is known as the broadcast problem [25]. To reduce this problem, improved flooding schemes have been proposed. In the original flooding operation, gossip-based routing protocol has been proposed as a method to determine whether RREQ messages can be transmitted or not, based on a predetermined probability *p* [26]. In Section 3, AODV implementing this processing is called gossip-based AODV, which is called GAODV. Though this method can reduce the energy consumption of the entire network, the setting of the probability *p* is difficult. This is because the appropriate value of *p* differs depending on the density and moving speed of nodes. In [27], the impact of various parameters such as node speed on the performance of a fixed probabilistic approach [26] has been shown. Particularly, mobility and pause times have a substantial impact on the reachability of data transmitted from the source node to the destination node. In [28], a probabilistic approach that dynamically adjusts the rebroadcasting probability according to the node density and node movement has been proposed. Although the node density can be estimated using the packet counter, it is necessary to set the time interval and *l* threshold for checking the packet counter appropriately according to the environment. In [29], a probabilistic approach that dynamically adjusts the rebroadcasting probability by considering the network density has been proposed. Simulation results have shown that the scheme can improve the saved broadcast up to 50% without affecting the reachability, even under high mobility and density conditions. However, because the average number of neighboring nodes is determined by the number of mobile nodes in the network and the area of the network, throughput performance affects them. In [30], a probabilistic approach that dynamically adjusts the rebroadcasting probability according to the number of neighboring nodes is presented. This method has been shown to outperform the results in [29]. However, it requires

complicated processing to determine the probability; moreover, the optimal value depends on the environment and should be evaluated by various models.

In [28–30], a method has been proposed in which each node periodically transmits hello packets to recognize the number of surrounding nodes and dynamically determines the probability p . Because the probability should be calculated and the average number of neighboring nodes should be assessed in advance, the energy consumption associated with these overheads is a concern, as in [19–21]. However, [27–30] mainly demonstrate the throughput and reachability performance, whereas the energy consumption performance is not discussed. Moreover, the effect of the dynamical and probabilistic approach on the lifetime of the entire network has not been clarified. Reference [28] has highlighted the necessity for a new strategy that can dynamically adjust the broadcast probability based on the current state of the node. In [16, 31, 32], the energy consumption of MPR nodes, which are the targets of flooding in OLSR routing, is reduced based on the same concept described above.

3. Proposed Scheme

We propose a simple flooding scheme to transmit RREQ based on only the remaining power of own node in the flooding process without the other information such as control packets and complex calculations to grasp the network status. This paper shows the case where the proposed method is applied to AODV [18] routing as an example. The proposed scheme is shown in Figure 1. The node that receives the RREQ packet generates a random variable X from 0 to 100 by itself. The value of X is compared with the ratio of the remaining energy of the node, Y (%). If the value of Y is greater than or equal to the value of X , flood the received RREQ packet to the adjacent node; otherwise, discard it. As a result, though a node with a large amount of remaining power tends to flood, the opposite is said for a node with a small amount of remaining power. The proposed scheme is widely applicable to flooding-based wireless routing in OLSR [16], WSN, DTN [17], etc. Moreover, as the proposed scheme modifies only the flooding process, it has little impact on other processes and can therefore be easily implemented to the previous methods.

4. Performance Evaluation

4.1. Setup and Definition. We evaluate the performance of the proposed scheme using a discrete-event network simulator, ns3 [33] (version 3.20). Figure 2 shows the basic simulation model.

After the source node (S) sends the packets, the relay nodes (R) use a routing scheme to transport the packets to the destination node (D). For example, Figure 2 shows an ad hoc network in an evacuation area in a disaster situation [1]. Important information transmitted from the victim S arrives at the gateway D connected to the internet via the rescuer R with an ad hoc terminal. As a wireless channel, the path loss, with an exponential coefficient, α is considered. According to the implementation of ns3 [33], some energy

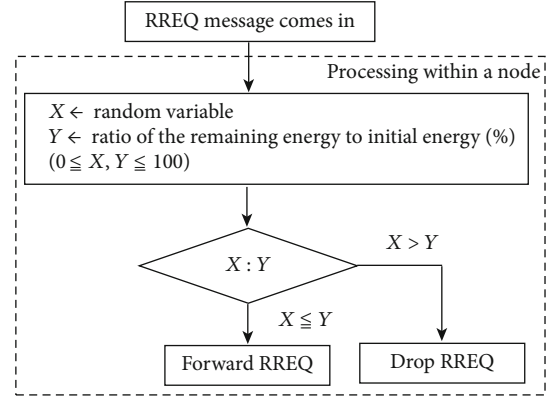


FIGURE 1: Proposed flooding scheme based on the remaining power of own node.

is consumed whenever an event occurs. The energy consumption of a node n is linearly approximated and expressed by

$$E_n(t + \Delta t_0) = E_n(t) + V \cdot \Delta t_0 \cdot I_i, \quad (1)$$

where $E_n(t)$ (J) is the energy consumption at time t (s), V (V) is the supply voltage, I_i (mA) is the current consumption defined for each event $i \in \{1, 2, \dots, 5\}$, and Δt_0 (s) is the time interval. The simulation parameters are presented in Table 1. The energy left in the node n is expressed by

$$\tilde{E}_n(t) = \beta \times E_0 - E_n(t), \quad (2)$$

where E_0 is the initial power of the battery (J), and β (%) is the ratio of the remaining energy to the full energy at the initial state. Hereafter, β is called the initial energy ratio.

We define five quantities to evaluate the proposed method. Throughput $TH(t)$ (Kbps) is the amount of data that arrives from the node S to the node D per unit time, expressed by

$$TH(T_{\text{sim}}) = \frac{P_D(T_{\text{sim}}) \cdot P_{\text{size}}}{T_{\text{sim}}}, \quad (3)$$

where $P_D(T_{\text{sim}})$ (packets) is the total number of packets received by the node D during one simulation time, T_{sim} (s), and P_{size} (Kbits/packet) is the packet size. $EC(T_{\text{sim}})$ (W) denotes the energy consumption per unit time; it is the total energy consumption in the whole network system during one simulation time, expressed by

$$EC(T_{\text{sim}}) = \frac{1}{T_{\text{sim}}} \sum_{n=1}^N E_n(T_{\text{sim}}), \quad (4)$$

where N is the number of all nodes. Energy efficiency EE (Kbits/J) is a combined measure of throughput and energy consumption. Using expressions (3) and (4), it is expressed by

$$EE(T_{\text{sim}}) = \frac{TH(T_{\text{sim}}) \cdot T_{\text{sim}}}{EC(T_{\text{sim}})}, \quad (5)$$

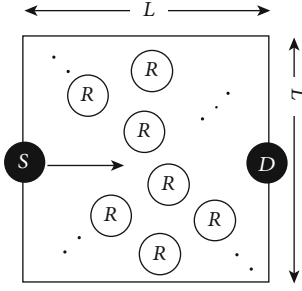


FIGURE 2: Simulation model. In a static scenario, the relay nodes R are static randomly. In a mobile scenario, the relay nodes R are moving randomly. In both scenarios, source node, S , and destination node, D , are fixed static.

where T_{sim} is one simulation time. The variance in remaining energy with all nodes, E_{σ} , at time, T_{sim} , is expressed by

$$E_{\sigma}(T_{\text{sim}}) = \frac{1}{N} \sum_{n=1}^N (\tilde{E}_n(T_{\text{sim}}) - E(\bar{T}_{\text{sim}}))^2, \quad (6)$$

where N is the total number of relay nodes, $\tilde{E}_n(T_{\text{sim}})$ is expressed by (2), and $E(\bar{T}_{\text{sim}})$ is the average remaining energy of all nodes at time, T_{sim} . The lifetime, LT (hour), is defined as the time until the ratio of total remaining power with all nodes becomes 10%. Table 1 shows the basic simulation parameters. Some routing-related parameters use request for comment (RFC) default values of AODV [18], and energy consumption parameters use the default values for ns3 (version 3.20). One simulation time, T_{sim} , is 300 seconds, and packet transmission from node S is suppressed for 20 seconds after the start of simulation to stabilize the routing operation. We consider two scenarios: (1) all relay nodes (R) are static, which is called the static scenario, (2) all relay nodes (R) are moving randomly, which is called the mobile scenario. Nodes S and D are always at a fixed position in these scenarios. In the static scenario, the random placement of nodes R is initialized in each simulation. In the mobile scenario, a random waypoint model defines the movement of node R . In this paper, we evaluated the performance of the proposed SAODV compared with the conventional schemes, GAODV and AODV as described in Section 2. In GAODV, the RREQ transmission probability, p , is set to 0.6 and 0.8 according to Reference [26].

4.2. Impact of the Initial Energy Ratio. We evaluated the case where the node density, ρ (nodes/km²), is 30. Each item is evaluated by uniformly increasing the ratio of the remaining energy, β from 10% to 100% in steps of 10% step for all nodes. Figure 3 shows the result of throughput, TH. In both static and mobility cases, when β is larger, all methods have the same performance, but when β is about 40% or less, the performance of the proposed method is degraded. Because the range that the RREQ message can reach is limited. For the entire range of β , the throughput of the mobile scenario is lower than that of the static scenario. This is because route disconnection occurs due to the node movement. Figure 4 shows the result of energy consumption, EC. Contrary to the throughput, energy consumption is reduced due to the

TABLE 1: Simulation parameters.

Communication standards	IEEE 802.11b
Simulation time (T_{sim})	300 sec
Side length of the target area (L)	1 km
Number of trials in one simulation	100
Number of relay nodes (N)	30
Velocity of relay node (v_r)	1.41 m/sec
Pass loss exponent (α)	3.5
Antenna gain	1
Limit coverage of direct transmission	250 m
Bit rate	1.0 Mbps
Application protocol	UDP
Traffic type	CBR
Generate traffic	2.0 Mbps
Packet size	1500 bytes
Routing protocol	AODV
Initialization energy of a node (E_0)	3600 J
Transmission power (P_t)	28 dBm
Receiver threshold	-96 dBm
Supply voltage (v)	3 V
Idle current (I_1)	0.426 mA
Clear channel assessment (CCA) busy state current (I_2)	0.426 mA
Channel switch current (I_3)	0.426 mA
Tx current (I_4)	17.4 mA
Rx current (I_5)	19.7 mA

operation of the proposed scheme. The energy consumption is almost constant in the conventional schemes, AODV, GAODV ($p = 0.6$), and GAODV ($p = 0.8$) regardless of the value of β . However, the energy consumption in the proposed scheme decreases as the value of β decreases in both static and mobile cases. Thus, a disadvantage related to throughput is an advantage with regard to power consumption. Figure 5 shows the result of energy efficiency, EE. Though all methods are almost equal in the static scenario, the proposed method has the best performance in the case of mobility.

Figure 6 shows the variance of the remaining energy for all nodes, E_{σ} . Although it is almost constant in the conventional schemes, it decreases as β decreases in the proposed scheme. In the proposed scheme, the remaining energy of all nodes can be equalized, and the battery replacement time is expected to remain approximately the same. This reduces the effort required for battery replacement.

4.3. Impact of Node Density. We evaluated the performances described in Section 4.1 for various node density ρ when β is 50%. Figure 7 shows the result of throughput, TH. In both static and mobile cases, when ρ is approximately 40 nodes/km² or more, the throughput of the proposed scheme becomes the highest. This contrasts with the result shown in Figure 3, where ρ is 30 nodes/km². Moreover, in a situation where ρ is higher, the throughput decreases as the RREQ

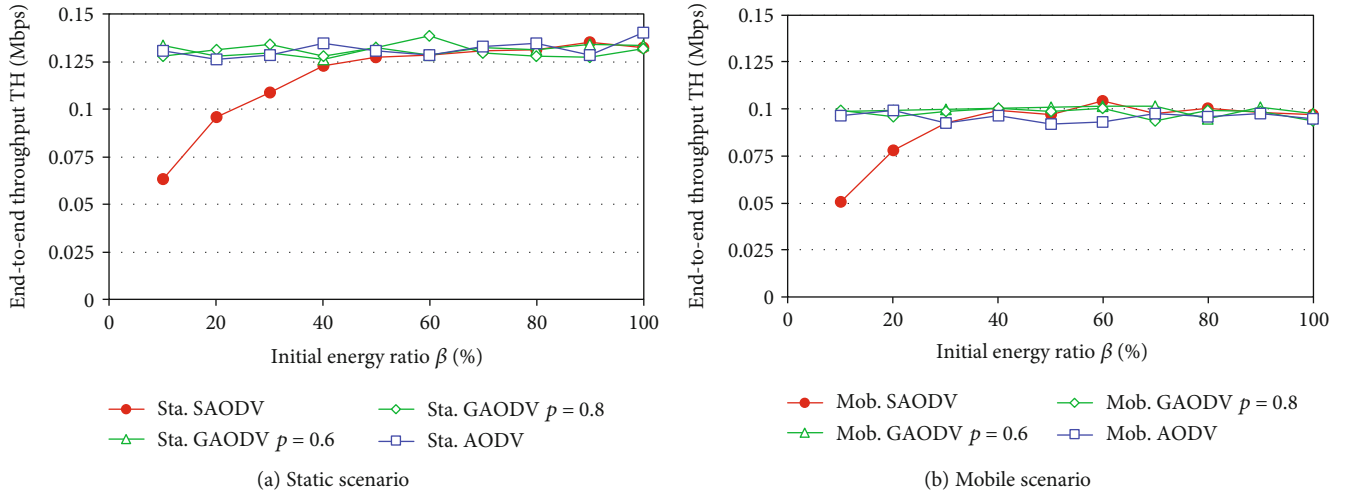


FIGURE 3: End-to-end throughput, TH with the initial energy ratio, β ($\rho = 30$ nodes/km²).

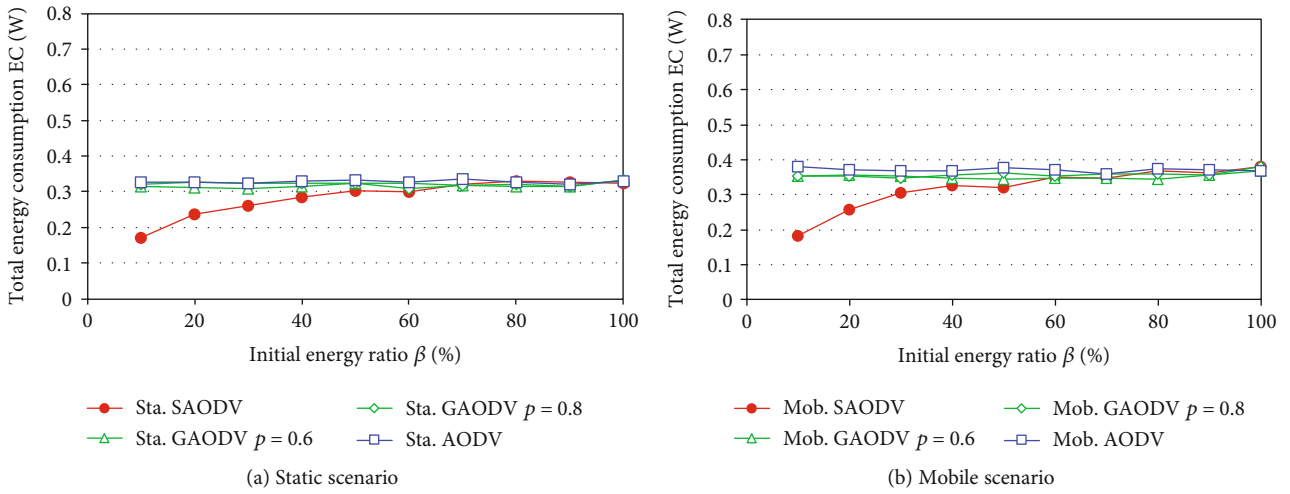


FIGURE 4: Energy consumption, EC, with the initial energy ratio, β ($\rho = 30$ nodes/km²).

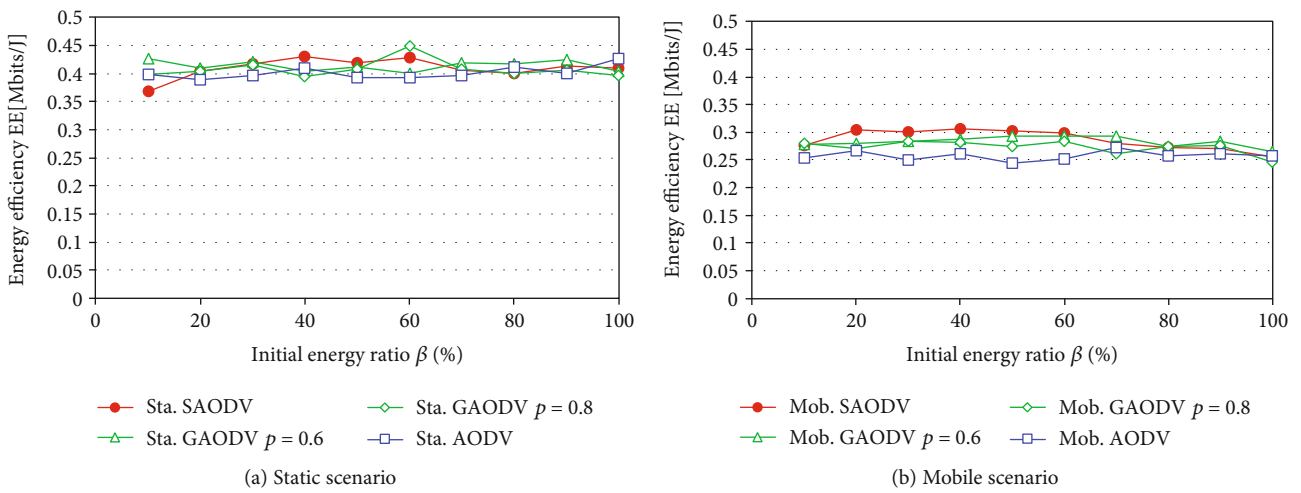


FIGURE 5: Energy efficiency, EE, with the initial energy ratio, β ($\rho = 30$ nodes/km²).

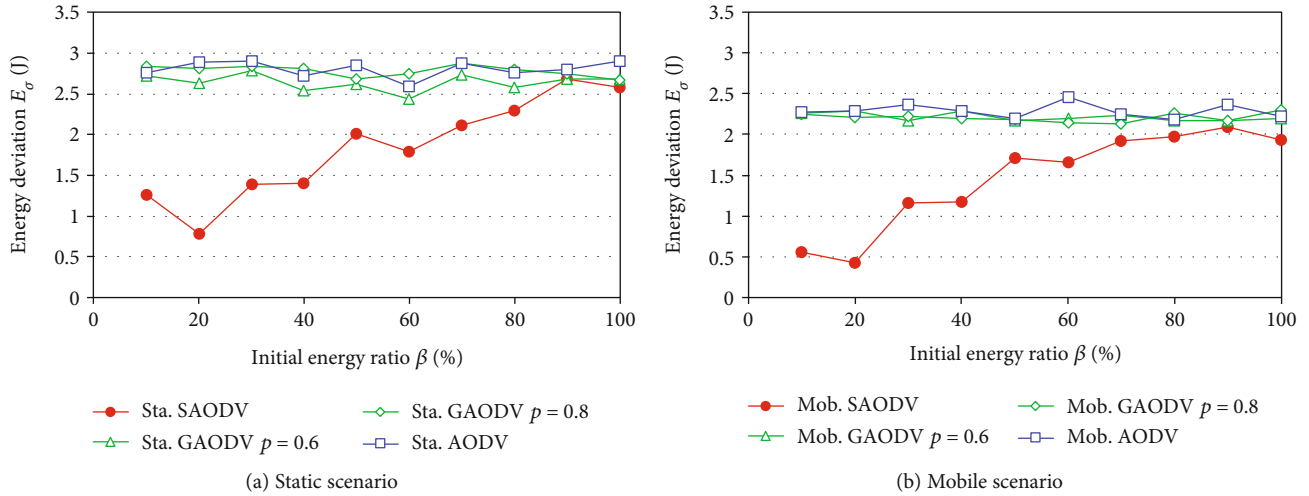


FIGURE 6: Variance of remaining energy, E_σ , with the initial energy ratio, β ($\rho = 30$ nodes/ km^2).

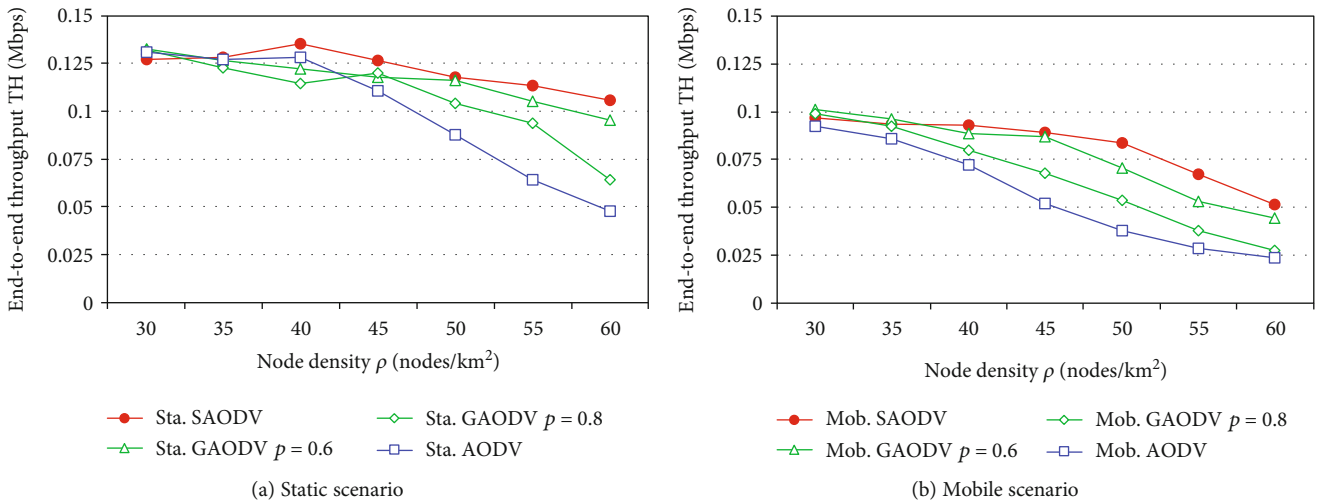


FIGURE 7: End-to-end throughput, TH, with node density, ρ ($\beta = 50\%$).

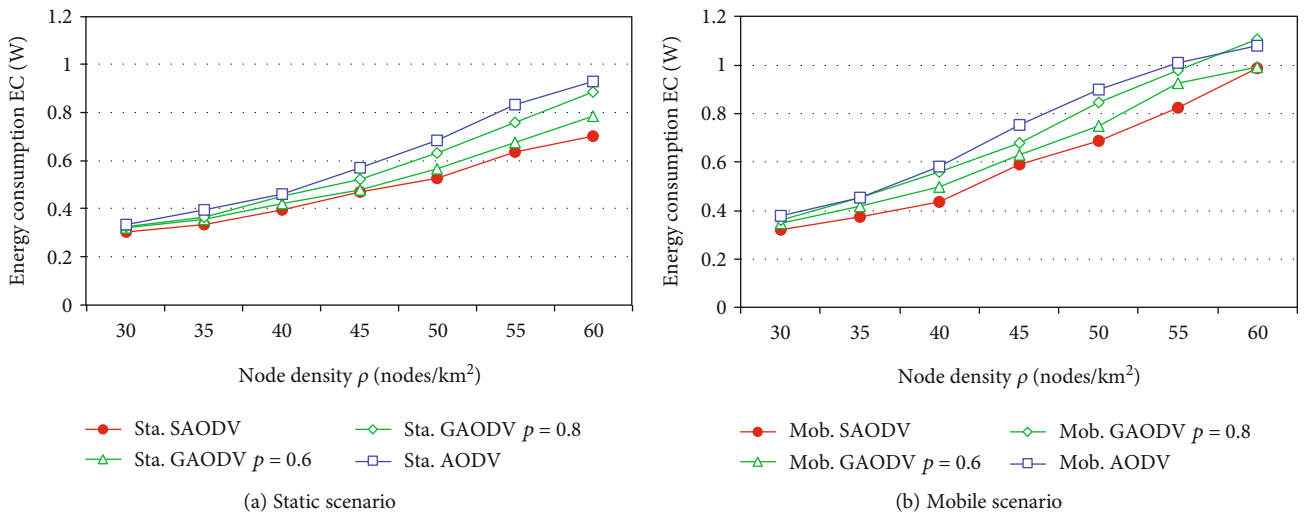


FIGURE 8: Energy consumption, EC, with node density, ρ ($\beta = 50\%$).

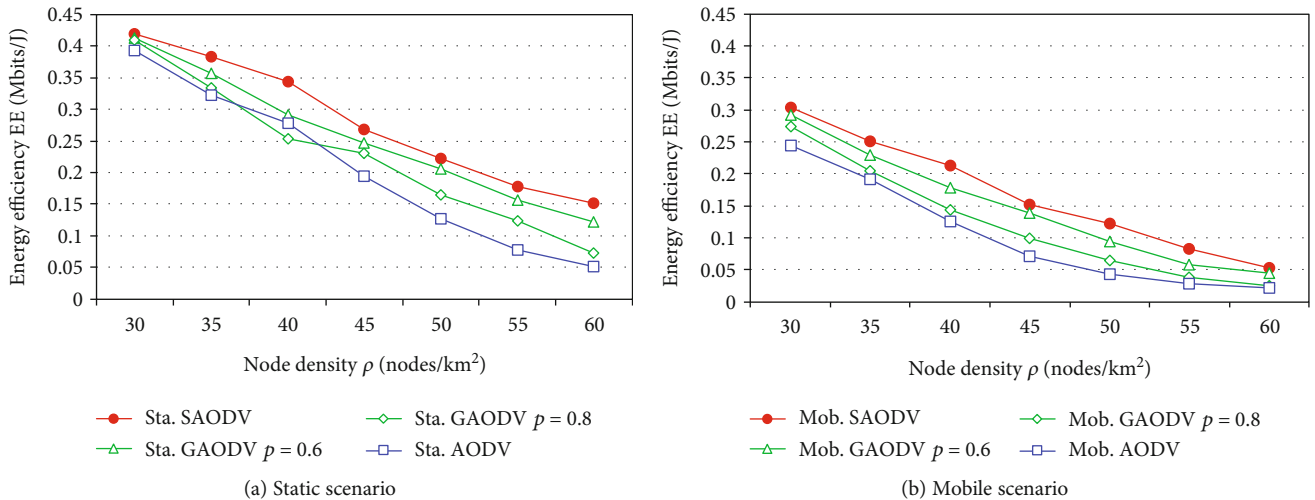


FIGURE 9: Energy efficiency, EE, with node density, ρ ($\beta = 50\%$).

transmission probability, p , increases, even in the conventional schemes. For these reasons, when ρ is larger, the ratio of RREQ messages required for route construction is larger than that required for information packets, and the throughput may be decreased. In [26], it is shown that almost every node gets the message using p between 0.6 and 0.8. Even in such a congested situation, as the proposed scheme suppresses the excessive transmission of RREQ messages, the avoidance of congestion and reduction of retransmitted packets can be achieved. As a result, higher throughput can be realized. On the other hand, in the static scenario, the best throughput is achieved when ρ is nearly 40; it should be noted that there exists an optimal ρ according to the environment in terms of throughput. Figure 8 shows the result of energy consumption. As shown in Figure 4, the proposed scheme reduces energy consumption due to its operation. Figure 8 shows that energy consumption decreases as node density increases. Figure 9 shows the result of energy efficiency, EE. From the results of throughput and energy consumption, we conclude that the energy efficiency in the proposed scheme is higher than conventional schemes regardless of the value of ρ .

Figure 10 shows the result for the lifetime, LT. In the static case, although the values of LT in conventional schemes decrease when ρ or p increases, it remains nearly the same in the proposed scheme. This is because in the conventional schemes, the energy consumption decreases LT, because of the increase in transmission and reception of RREQ messages. However, as the number of flooding times is made uniform in the proposed scheme, all nodes have almost an equal chance to transmit the RREQ messages. This consideration is also related to the results in Figure 6. The values of LT are smaller in the case of movement than those in the static scenario. It is considered that energy is consumed by transmitting and receiving control packets to retransmit when a communication path is disconnected due to node movement. However, even in the case of movement, the proposed scheme has better performance than conventional schemes. In particular, we showed that the lifetime

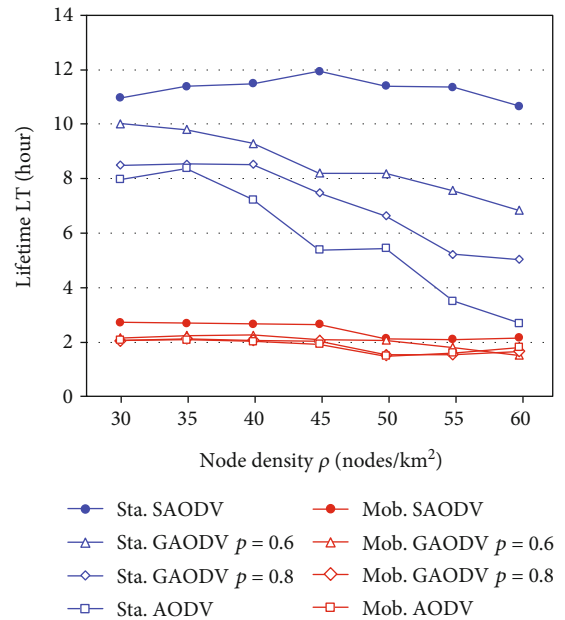


FIGURE 10: Lifetime, LT, with node density, ρ ($\beta = 50\%$).

was extended up to about 2.2 times in the static scenario and about 1.4 times in the mobile scenario compared to the original AODV when the node density was 45 nodes/km².

4.4. Impact of the Number of Flows. We discussed the cases in which the number of flows, in other words, number of node S was one shown in Figure 1. In this section, we evaluate by increasing the number of flows to confirm the further effects of the proposed method. The number of source nodes S is increased at equally spaced positions on the left-hand line segment with one destination node D fixed, as shown in Figure 2. Figure 11 shows the impact of LT on the number of flows, F . The consideration for the result in Figure 11 is almost equal to that of Figure Communication standards IEEE 802.11b Simulation time (T_{sim}) 300 sec Side length of

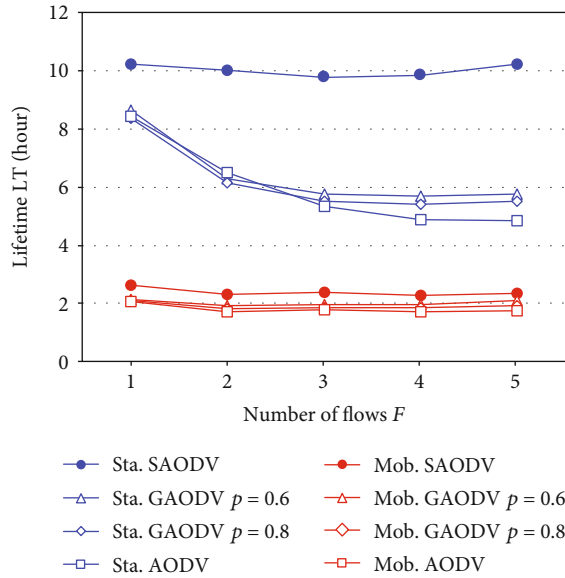


FIGURE 11: Lifetime, LT, with the number of flows, F ($\beta = 50\%$).

the target area (L) 1 km Number of trials in one simulation 100 Number of relay nodes (N) 30 Velocity of relay node (v_r) 1.41 m/sec Pass loss exponent (α) 3.5 Antenna gain 1 Limit coverage of direct transmission 250 m Bit rate 1.0 Mbps Application protocol UDP Traffic type CBR Generate traffic 2.0 Mbps Packet size 1500 bytes Routing protocol AODV Initialization energy of a node (E_0) 3600 J Transmission power (P_t) 28 dBm Receiver threshold -96 dBm Supply voltage (v) 3 V Idle current (I_1) 0.426 mA Clear channel assessment (CCA) busy state current (I_2) 0.426 mA Channel switch current (I_3) 0.426 mA Tx current (I_4) 17.4 mA Rx current (I_5) 19.7 mA . The proposed scheme can keep the lifetime constant regardless of the number of flows. This is because increasing the node density and the number of flows can be considered to have equivalent effects. In particular, we showed that the lifetime extended up to nearly 2.0 times in the static scenario and 1.3 times in the mobile scenario compared to the original AODV when the number of flows was 3. Thus, the proposed scheme is more effective, even in an environment with a larger number of flows.

5. Conclusions

We proposed a simple flooding scheme to transmit RREQ message based on only the remaining power of each node in the flooding process without additional information such as control packets and complex calculations. We applied the proposed scheme to AODV of MANET routing and showed the superiority of the proposed scheme to the conventional scheme using computer simulation. As the node density decreases, the throughput of the proposed scheme becomes lower than that of the conventional methods, whereas the throughput of the proposed scheme increases when the node density becomes higher than nearly 40 in the evaluated environment. Regardless of the node density, energy consumption was reduced in the proposed scheme.

As a result, we showed that the proposed scheme was superior to conventional schemes in terms of energy efficiency (bits/J), including throughput and energy consumption. Moreover, as the nodes to be flooded are equalized in the proposed scheme, the variance in the remaining energy of all nodes becomes smaller. The proposed approach ensures that all nodes need almost the same battery replacement time. Thus, the lifetime in the proposed scheme does not decrease when the nodes are static, regardless of the node density or the number of flows.

Future works include the evaluation of the proposed scheme for unmanned aerial vehicle (UAV) networks [34] and comparison with other dynamical and probabilistic approaches.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

This work was in part supported by the 2019 Telecommunications Advancement Foundation.

References

- [1] H. Nishiyama, M. Ito, and N. Kato, "Relay-by-smartphone: realizing multihop device-to-device communications," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 56–65, 2014.
- [2] S. Corson and J. Macker, "Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations," RFC 2501, 1999.
- [3] C. E. Perkins, *Ad Hoc Networking*, Addison-Wesley, 2001.
- [4] S. Ci, M. Guizani, H. H. Chen, and H. Sharif, "Self-regulating network utilization in mobile ad hoc wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1302–1310, 2006.
- [5] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213)*, vol. 3, pp. 1548–1557, Anchorage, AK, USA, 2001.
- [6] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, no. 4, pp. 8–27, 2002.
- [7] C. Yu, B. Lee, and H. Y. Youn, "Energy efficient routing protocols for mobile ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 3, no. 8, pp. 959–973, 2003.
- [8] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings-energy optimization and next-generation building management systems," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 269–283, 2017.

- [9] A. Batra, A. Shukla, S. Thakur, and R. Majumdar, "Survey of routing protocols for mobile ad hoc networks," *IOSR Journal of Computer Engineering*, vol. 8, no. 1, pp. 34–40, 2012.
- [10] S. Prakash, J. Saini, and S. Gupta, "A review of energy efficient routing protocols for mobile ad hoc wireless networks," *International Journal of Computer Information Systems*, vol. 1, no. 4, pp. 36–46, 2010.
- [11] I. Demirkol, C. Ersoy, and F. Alagöz, "MAC protocols for wireless sensor networks: a survey," *IEEE Communications Magazine*, vol. 44, no. 4, pp. 115–121, 2006.
- [12] V. KumarSachan, S. A. Imam, and M. T. Beg, "Energy-efficient communication methods in wireless sensor networks: a critical review," *International Journal of Computer Applications*, vol. 39, no. 17, pp. 35–48, 2012.
- [13] Z. Ul, A. Jaffri, and S. Rauf, "A survey on "energy efficient routing techniques in wireless sensor networks focusing on hierarchical network routing protocols"," *International Journal of Scientific and Research Publications*, vol. 4, no. 1, pp. 2250–3153, 2014.
- [14] S. Cui, R. Madan, A. Goldsmith, and S. Lall, "Joint routing, MAC, and link layer optimization in sensor networks with energy constraints," in *IEEE International Conference on Communications, 2005*, vol. 2, pp. 725–729, Seoul, South Korea, 2005.
- [15] F. Adachi, "Wireless network evolution toward 5G network," in *2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS)*, pp. 1–3, Niigata, Japan, 2016.
- [16] T. Clausen and P. Jacquet, Eds., "Optimized link state routing protocol (OLSR)," RFC 3626, 2003.
- [17] V. Cerf, S. Burleigh, A. Hooke et al., "Delay-tolerant networking architecture," RFC 4838, 2007.
- [18] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," RFC 3561, 2003.
- [19] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocols for wireless microsensor networks," in *Proceedings of the 33rd Hawaiiian International Conference on Systems Science (HICSS)*, pp. 1–10, Maui, HI, USA, 2000.
- [20] R. C. Shah and J. M. Rabaey, "Energy aware routing for low energy ad hoc sensor networks," in *2002 IEEE Wireless Communications and Networking Conference Record*, vol. 1, pp. 350–355, Orlando, FL, USA, 2002.
- [21] J. Ben-Othman and B. Yahya, "Energy efficient and QoS based routing protocol for wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 70, no. 8, pp. 849–857, 2010.
- [22] X. Jing and M. J. Lee, "Energy-aware algorithms for AODV in ad hoc networks," in *Proceedings of Mobile Computing and Ubiquitous Networking*, pp. 466–468, Yokosuka, Japan, 2004.
- [23] S. P. Bhatsangave and V. R. Chirchi, "OAODV routing algorithm for improving energy efficiency in MANET," *International Journal of Computer Applications*, vol. 51, no. 21, pp. 15–22, 2012.
- [24] D. Johnson, Y. Hu, and D. Maltz, "The dynamic source routing protocol (DSR) for mobile ad hoc networks for IPv4," RFC 4728, 2007.
- [25] Y. C. Tseng, S. Y. Ni, Y. S. Chen, and J. P. Sheu, "The broadcast storm problem in a mobile ad hoc network," *Wireless Networks*, vol. 8, no. 2/3, pp. 153–167, 2002.
- [26] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-based ad hoc routing," in *Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3no. 3, pp. 1707–1716, New York, USA, 2002.
- [27] M. Bani-Yassein, M. Ould-Khaoua, L. M. MacKenzie, and S. Papanastasiou, "Performance analysis of adjusted probabilistic broadcasting in mobile ad hoc networks," *International Journal of Wireless Information Networks*, vol. 13, no. 2, pp. 127–140, 2006.
- [28] Q. Zhang and D. P. Agrawal, "Dynamic probabilistic broadcasting in MANETs," *Journal of Parallel and Distributed Computing*, vol. 65, no. 2, pp. 220–233, 2005.
- [29] M. B. Yassein, M. O. Khaoua, L. M. Mackenzie, S. Papanastasiou, and A. Jamal, "Improving route discovery in on-demand routing protocols using local topology information in MANETs," in *PM2HW2N 2006: Proceedings of 2006 ACM International Workshop on Performance Monitoring, Measurement, and Evaluation of Heterogeneous Wireless and Wired Networks 2006*, pp. 95–99, Terromolinos, Spain, October 2014.
- [30] A. M. Hanashi, A. Siddique, I. Awan, and M. Woodward, "Performance evaluation of dynamic probabilistic broadcasting for flooding in mobile ad hoc networks," *Simulation Modelling Practice and Theory*, vol. 17, no. 2, pp. 364–375, 2009.
- [31] T. Kunz, "Energy-efficient variations of OLSR," in *2008 International Wireless Communications and Mobile Computing Conference*, pp. 517–522, Crete Island, Greece, September 2008.
- [32] H. Chizari, M. Hosseini, S. Salleh, S. A. Razak, and A. H. Abdullah, "EF-MPR, a new energy efficient multi-point relay selection algorithm for MANET," *Journal of Supercomputing*, vol. 59, no. 2, pp. 744–761, 2012.
- [33] Anon, "ns-3, a discrete-event network simulator," 2008.
- [34] T. Li, W. Liu, T. Wang, Z. Ming, X. Li, and M. Ma, "Trust data collections via vehicles joint with unmanned aerial vehicles in the smart internet of things," *Transactions on Emerging Telecommunications Technologies*, vol. 2020, pp. 13–16, 2020.

Research Article

Reducing the Energy Budget in WSN Using Time Series Models

Felicia Engmann ^{1,2}, **Ferdinand Apietu Katsriku**,² **Jamal-Deen Abdulai**,²
and **Kofi Sarpong Adu-Manu**²

¹*School of Technology, GIMPA, Accra, Ghana*

²*Department of Computer Science, University of Ghana, Legon, Accra, Ghana*

Correspondence should be addressed to Felicia Engmann; fnaengmann@st.ug.edu.gh

Received 21 March 2020; Revised 23 June 2020; Accepted 7 July 2020; Published 3 August 2020

Academic Editor: Yujin Lim

Copyright © 2020 Felicia Engmann et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Energy conservation is critical in the design of wireless sensor networks since it determines its lifetime. Reducing the frequency of transmission is one way of reducing the cost, but it must not tamper with the reliability of the data received at the sink. In this paper, duty cycling and data-driven approaches have been used together to influence the prediction approach used in reducing data transmission. While duty cycling ensures nodes that are inactive for longer periods to save energy, the data-driven approach ensures features of the data that are used in predicting the data that the network needs during such inactive periods. Using the grey series model, a modified rolling GM(1,1) is proposed to improve the prediction accuracy of the model. Simulations suggest a 150% energy savings while not compromising on the reliability of the data received.

1. Introduction

Wireless sensor network (WSN) is the backbone of ubiquitous computing applications such as military surveillance, disaster recoveries, environmental and structural monitoring, health and security monitoring and control, wildlife monitoring and precision agriculture, and habitat monitoring. Sensor nodes which are the basic components in a WSN may sense the environment, aggregate and compute the data collected, and usually transmit the data from source nodes to a sink node called a base station [1].

With the growing need to monitor and control the environment, wireless sensor networks have become an indispensable tool enabling the required data to be acquired and transmitted to relevant endpoints. They are based on the microelectromechanical (MEM) technology, radio or wireless technology, and digital electronics to enable low-cost low-powered miniature devices to communicate untethered over near and far distances. WSNs are made up primarily of sensor nodes, otherwise known as motes which have sensing, data processing, and communicating capabilities and are usually deployed in large numbers due to their

size and low cost. Wireless sensor networks may consist of a base station and/or a gateway and several deployed sensor nodes whose positions may be randomly determined or predetermined.

Characteristics of wireless sensor networks include but not limited to their miniature sizes and the low power. They, however, do have some drawbacks such as limited memory space, low computational power, and the low bandwidth [2].

The miniaturized nodes influence the number of sensor nodes mostly deployed in applications, mostly ranging between several hundreds to thousands. It also limits the number of components on the microchip, limiting their capabilities. This suggests that the nodes usually have low power capability. The power constraint in WSN is a major consideration in their deployment and use. Of the major components (radio, processor, storage, and power unit) of the typical node, the radio which supports communication is the major consumer of the energy of sensor nodes. The energy expended by the radio may account for up to 70% of the total energy expended in the sensor node [3].

The sensor networks may either be deployed as star, tree, mesh, or point-to-point as shown in Figure 1.

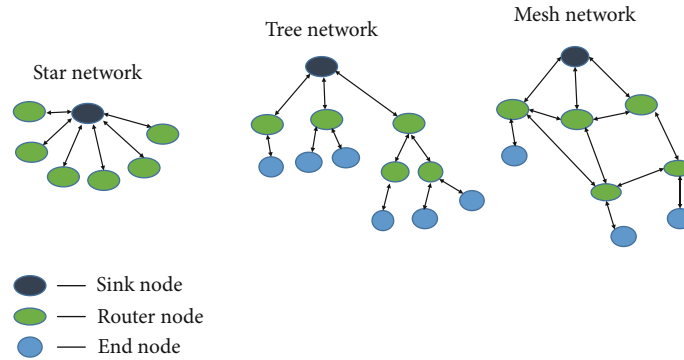


FIGURE 1: Topology of wireless sensor networks.

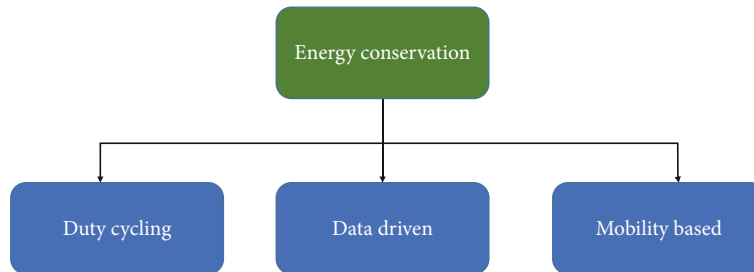


FIGURE 2: Energy conservation in wireless sensor networks [1].

The mesh topology is mostly employed in wireless sensor networks to mitigate the problem of energy consumption in WSN discussed above. This is because mesh topologies have low transmit power, transmitting over shorter distances as opposed to other topologies that allow long single-hop transmissions to the sink. In WSN applications where continuous monitoring of sensed phenomenon is extremely important (e.g., health monitoring systems, structural health monitoring systems, road traffic monitoring systems, and water quality monitoring systems [4]), continuous communication of the sensed data is essential to maintain the reliability of the data received at the base station and importantly to be able to detect any changes in the sensed environment. Such continuous transmissions add communication costs of the network that may deplete the energy of the batteries powering the nodes. The resultant effect is network failure due to several nodes dying off before the end of the network's operational lifetime [5]. To ensure the reliability of data received at the base station, routing, duty cycling and retransmissions, and redundancies are some approaches used. Lessening the number of retransmissions will reduce the amount of energy wasted since dropped data packets are not transmitted again while introducing redundant nodes increases the overall energy cost of transmission in the network.

The implementation of WSN in environmental monitoring applications usually requires uninterrupted operations. This is to ensure that no incident is missed, and the data received can be relied upon. To guarantee the reliable acquisition and transmission of data, the power supplied to the node must be maintained or replenished [6]. However, continuous monitoring and transmission lead to power depletion. In recent times, there has been an increased interest in

the use of predictive algorithms to augment the acquisition and transmission of actual data.

In this paper, we analyze two time series models, the autoregressive integrated moving average (ARIMA) and grey model (GM(1,1)) as approaches to model and forecast time series data in WSN. The prediction modeling used is aimed at determining the optimal period for scheduling the *on* and *off* periods of nodes in a network. Nodes will be scheduled to be turned *off* for the duration of an optimum period determined by the prediction algorithm and turned *on* when new data must be collected for processing. During the period, when the node is turned off, the data for that period is predicted. Our simulations indicate that we were able to determine the optimum number of data sets required and the longest sleep period required for the prediction that minimizes energy consumption without compromising on the reliability and accuracy of data received.

2. Literature Review

2.1. Conserving Energy in the Network. Proposed energy conservation techniques in the literature are broadly categorized as duty cycling, data-driven, and mobility [7, 8] as shown in Figure 2. These approaches have been studied extensively in the literature either individually or in combination to mitigate energy consumption in the wireless sensor networks. Duty cycling approaches put the radio in low power or sleep mode for longer periods when there is no data to be transmitted. The nodes, therefore, alternate between sleep and active modes due to the activities of the sensor node. A duty cycle may be defined as the ratio of the active time to the sleeping time [9, 10]. Since duty cycling is generally not concerned

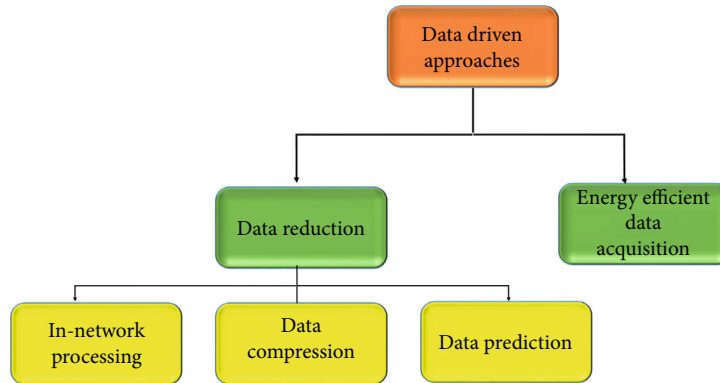


FIGURE 3: Data-driven approached in wireless sensor networks [1].

about the characteristics of data in transmission [7], data-driven approaches use the features of the data in transition to mitigate energy consumption.

Duty cycling approaches determine optimum schedules to turn *off* or turn *on* nodes for transmission with the aim of reducing energy consumption in the network. The aim is to reduce idle listening when the node's radio waits in vain for frames and overhearing when nodes stay active listening to uninterested frames. Duty cycling approaches usually include scheduling algorithms that determine periods when nodes transition between sleep and wake up modes. The active periods of the individual sensor nodes usually have the same length called the slot time. The transceiver of sensor nodes may transmit packets at any time, but can only receive packets when in the active state. Duty cycling reduces idle listening, overhearing, and overlistening, which are major causes of energy loss in WSN [10]. The scheduling algorithms implemented in duty cycling eliminate idle listening since it gives nodes their shared access to the wireless medium [11]. Scheduling algorithms reduce idle listening and also ensure the efficient transmission of data between neighboring nodes that may be in sleep mode and nodes in active mode. While the duty cycle is fixed and determined before network deployment, scheduling algorithms may be used in addition to adjusting the sleep-wake cycle to the network traffic.

Data-driven approaches are aimed at reducing the sampling and transmission of redundant data even if sensing, as in some instances, constitutes only an insignificant portion of the energy consumption [12, 13]. It also includes the reduction of the amount of data transmitted without negatively affecting the quality and reliability of the resultant data [14, 15]. In applications such as real-time systems where continuous monitoring is essential, data-driven approaches ensure that energy consumed by sensing is maintained within allowable thresholds [16, 17].

Mobility may be an important consideration in the energy conservation scheme of sensor networks [18]. Mobility ensures that the physical location of nodes or the sink is adjusted to primarily reduce the distance between nodes to minimize energy consumption. It therefore mitigates the problems of static nodes during multihop communication, which usually have the energy hole problem where nodes closer to the sink are depleted of their energy [19].

Conservation of energy is usually achieved through the use of two or more energy conservation approaches. In this paper, data-driven approaches are used in conjunction with data reduction to reduce the frequency of transmission.

Data prediction as a data reduction method in WSNs is concerned with the building of a model for forecasting sensed parameters of sensor nodes using historical data [17]. The forecasted values are adopted when they fall within some predefined acceptable thresholds. Applications that use prediction in WSNs have the prediction model residing on the sink, the sensor node, or mostly on both. In most prediction approaches, the trend of data received at a node is modeled, such that only data deviating from the trend is communicated to the sink. This approach requires periodic synchronization of the model on the nodes with that on the sinks when received data deviates from the forecasted values. Data prediction approaches in the literature have been categorized into three main areas, which are stochastic, time series, and algorithmic approaches [7] as shown in Figure 3. Recent research has introduced machine learning algorithms into WSN with some successful applications [20–22]. Time series models have been used to predict future data values in WSN mostly to reduce the frequency of data transmission. Models such as autoregressive integrated moving average (ARIMA), autoregressive moving average (ARMA), least mean squares (LMS), and grey series have also been used in WSN.

The ARIMA models both stationary and stochastic data by “integrating” the simpler autoregressive (AR) and moving average (MA) models [23]. The ARIMA is based on the Box-Jenkins methodology [24]. The grey series models [25] stochastic data whose values vary with time and its advantageous in forecasting time series data since the modeling is based on the generated series in the modeling process and not the raw data. It also supports the near accurate modeling of data as small as 4 data sets. Time series models are increasingly used in WSN, with the most common ones being ARIMA [23, 26], least mean squares [27, 28], and the grey series [29, 30]. In this paper, the ARIMA and grey series are compared for best performance.

2.2. Data Prediction Approaches. In WSN, time series models are increasingly being used to predict and reduce data transmission. Autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated

moving average, and least mean square algorithm are some linear algorithms used to WSN. Their primary advantages over other predictive techniques are that they exhibit the least complexities; they are simple to implement and produce acceptable accuracies [31]. The accuracy of a time series model depends on the nature of the data sampled. For example, ARMA models are suitable for slowly changing data like the temperature of a water body and its water level. Data that may exhibit sudden or sharp changes such as potential of hydrogen (pH) which may be a result of unforeseen human factors are best modeled with MA and give a better ratio of their complexity and performance. Grey series are time series models that are known for their superiority over other time series models due to their high accuracy and simplicity of use. The GM(1,1), also known as “grey series one differential one variable”, is the most common of the grey series models used [32, 33]. It is usually used to predict both linear and nonlinear data series. In this work, two popularly used time series models employed in WSN are discussed, and simulation analysis is performed to select the optimum energy-efficient method without compromising on reliability.

2.2.1. ARIMA. An ARIMA model is of order ARIMA(p, q, r), where p is the number of autoregressive terms, q is the non-seasonal difference needed for stationary data, and r is the number of lagged forecast errors in the prediction. The prediction is such that

$$\begin{aligned} \text{If } q = 0 : x_t &= X_t, \\ \text{If } q = 1 : x_t &= X_t - X_{t-1}, \\ \text{If } q = 2 : x_t &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned} \quad (1)$$

\forall values of X time series data are described above.

ARIMA may be applied to stationary data, which means the data series may have no trend, with little variations of the mean that has a constant amplitude and has its short-term random patterns looking the same over time. The reader may refer to [26, 34] for detailed explanation on using ARIMA for WSN data prediction.

2.2.2. Grey Series. The grey series was initially presented by Deng in 1982 [35, 36] and has been applied several in finance, physical control, engineering, and economics. It is an example of a time series prediction model that shows superiority to other statistical methods of prediction [25]. Artificial intelligent models in the literature include fuzzy systems [37], hidden Markov models [38], support vector machines [39], and neural networks [22, 40]. The intelligent approaches such as the neural networks give better accuracies in prediction than their statistical counterparts but are usually very complex to implement. Neural networks (NN) are also criticized as requiring a higher number of training data set that need longer periods to train before generalizations are acceptable. The grey series approach, also known as the GM(n, m), is grey model, where n is the order of the differen-

tial equation and m is the number of variables. It is widely used in the prediction of time-dependent data because of its ability to predict data that has uncertain characteristics while avoiding conventional statistical properties. Its advantages include but not limited to its ability to discover inherent regularity of data from disorganized data to predict its performance variations at some point in the future [36].

Steps to the grey series include the following:

(i) Original data:

$$\text{Let } X(0) = (x(0)(1), x(0)(2), \dots, x(0)(n)) \quad (2)$$

be a time series data obtained from the environment and n is the period when data was collected with the difference between subsequent values being some time t .

(ii) The accumulation generation sequence $X^{(1)}$ is obtained from $X^{(0)}$ by

$$\text{AGO}\{x^{(0)}(k)\} = x^{(1)}(k) \left(\sum_{k=1}^1 x^{(0)}(k), \sum_{k=1}^2 x^{(0)}(k), \dots, \sum_{k=1}^n x^{(0)}(k) \right), \quad (3)$$

where $x^{(1)}(k) = \sum x^{(0)}(i), k = 1, 2, 3, \dots, n$.

(iii) The original form of the GM(1,1) is found in the following equation:

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (4)$$

Using the adjacent neighbor means, the generated sequence is obtained from $X^{(0)}$ to obtain the sequence in the following equation.

$$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)). \quad (5)$$

And the basic form of the GM(1,1) is shown in the following equation.

$$x^{(0)}(k) + az^{(1)}(k) = b. \quad (6)$$

For a series of data, $X(k) = x(1), x(2), x(3), \dots, x(n)$, the value of $Z(k) = 1/2(x(k) + x(k-1))$. The sequence of data obtained by the adjacent neighbor means constructs the new series for prediction based on the available time series data.

(iv) Using the least squared means to obtain the parameters of the GM(1,1), $x^{(0)}(k) + az^{(1)}(k) = b$, putting in the values for each k

$$\begin{aligned}
x^{(0)}(2) &= -az^{(1)}(2) + b, \\
x^{(0)}(3) &= -az^{(1)}(3) + b, \\
x^{(0)}(4) &= -az^{(1)}(4) + b, \\
&\dots \\
x^{(0)}(n) &= -az^{(1)}(n) + b.
\end{aligned} \tag{7}$$

Convert the above equation into a matrix of the form $Y = Ba^*$

$$\begin{bmatrix} x^{(0)} & 2 \\ x^{(0)} & 3 \\ x^{(0)} & 4 \\ \vdots & \vdots \\ x^{(0)} & 1n \end{bmatrix} = \begin{bmatrix} z^{(1)}(2) & 1 \\ z^{(1)}(3) & 1 \\ z^{(1)}(4) & 1 \\ \vdots & \vdots \\ z^{(1)}(n) & 1 \end{bmatrix} * \begin{bmatrix} a \\ b \end{bmatrix}, \tag{8}$$

where

$$\begin{aligned}
Y &= \begin{bmatrix} x^{(0)} & 2 \\ x^{(0)} & 3 \\ x^{(0)} & 4 \\ \vdots & \vdots \\ x^{(0)} & 1(n) \end{bmatrix}, \\
B &= \begin{bmatrix} z^{(2)}(2) & 1 \\ z^{(2)}(3) & 1 \\ z^{(2)}(4) & 1 \\ \vdots & \vdots \\ z^{(2)} & 1(n) \end{bmatrix}, \\
\hat{a} &= \begin{bmatrix} a \\ b \end{bmatrix}.
\end{aligned} \tag{9}$$

(v) Using $\hat{a} = (B^T B)^{-1} B^T$, solve for the values of a and b and the values of

$$\begin{aligned}
a &= \frac{CD - (n-1)E}{(n-1)F - C^2}, \\
b &= \frac{DF - CE}{(n-1)F - C^2},
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
C &= \sum_{k=2}^n z^{(1)}(k), \\
D &= \sum_{k=2}^n x^{(0)}(k), \\
E &= \sum_{k=2}^n z^{(1)}(k)x^{(0)}(k), \\
F &= \sum_{k=2}^n z^{(1)}(k)^2,
\end{aligned} \tag{11}$$

where n is the number of data values. The parameters a and b are known as the development coefficient and grey action quantity of the GM(1,1). GM(1,1) being a univariate sequence depends on the backgrounds values a and b . The value of b reflects the changes in the data sequence and its intention is grey as opposed to other black box or white box methods. The model is invalid when $|a| \geq 2$, such that for values of $(-\infty, -2) \cup (\infty, 2)$ the growth rate weakens and the simulation error of the prediction grows drastically.

The predicted values are obtained from Equation (13), which is the expanded form of the following equation.

$$\hat{x}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}. \tag{12}$$

Reducing the inverse AGO, the discrete response of the GM(1,1) is

$$x^{\wedge(0)}(k+1) = (1 - e^a) \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak}, \tag{13}$$

which gives the final prediction value obtained from the even difference grey model.

3. Results and Discussion

Given a continuously flowing river, the quality parameters at a position (x_1, y_1) at time t_n may differ significantly from parameters acquired at the same position at a different time, t_{n+1} . The quality parameters are therefore temporal, attributable to the flow of the river and changes in the river ecosystem. These changes may result from introduction pollutants into the water or through other human activities. The temperature of the water is also subject to changes, varying with the time of day and the incidence of sunshine on the river.

The accuracy of several predictive models is largely dependent on the size of the data. Large data sets generally give better predictions than fewer data sets. Original readings of water quality parameters from a river are collected and presented as shown in Figure 4. 48-hour readings of conductivity, temperature, and pH taken from the 12:00 am on November 1, 2018, to 12:00 am November 3, 2018 are presented. The total number of data points obtained in the 48 hours was 88 for each of the parameters, pH, conductivity,

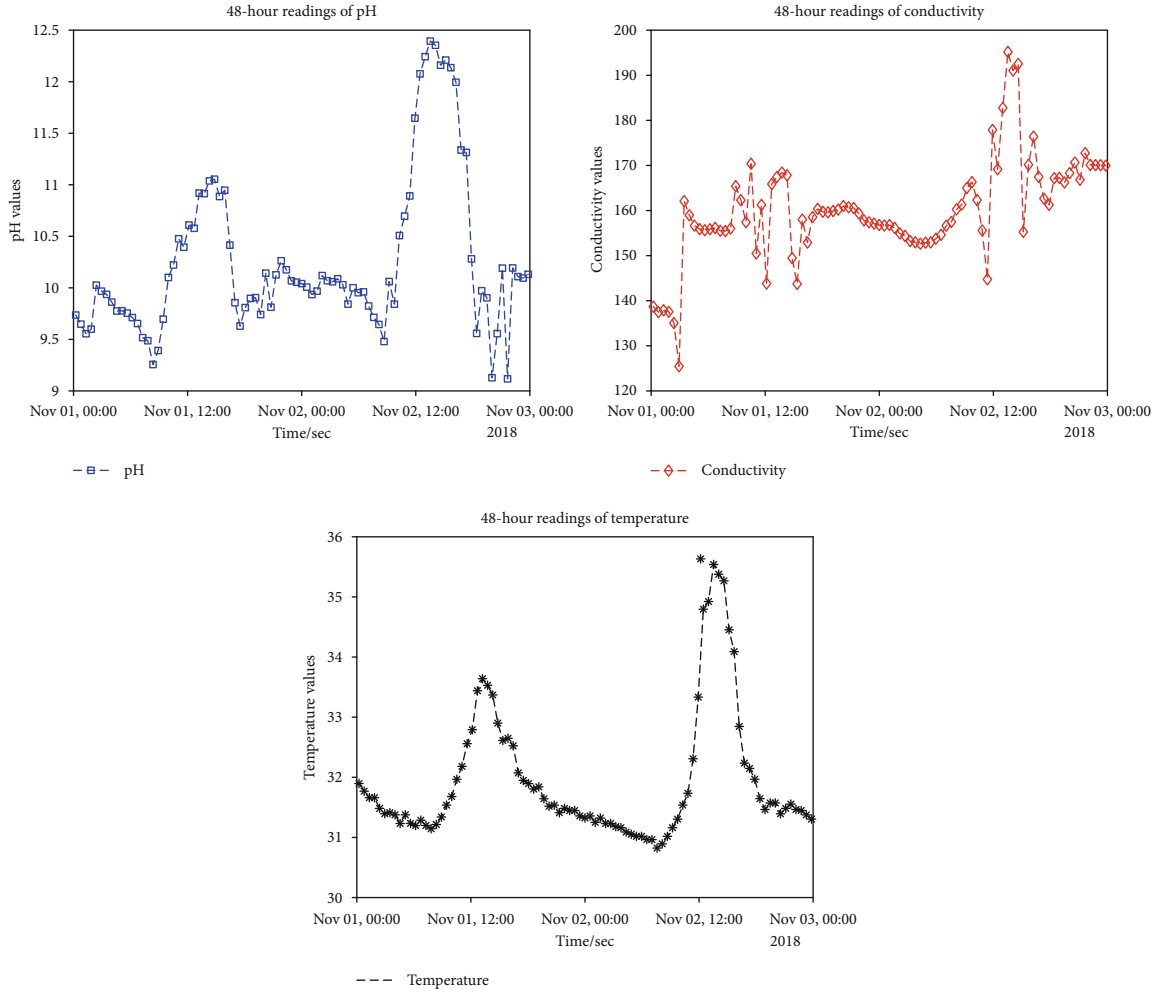


FIGURE 4: 48-hour readings of pH, conductivity, and temperature sensors.

and temperature. The best models of ARIMA for 15, 25, and 35 readings were ARIMA(1,0,0), ARIMA(0,1,0), and ARIMA(0,1,0), respectively.

Conductivity values for GM(1,1) and ARIMA were predicted with varying data points as a training set and the remaining as the test set for the model. The models obtained were ARIMA(0,1,0), ARIMA(0,1,1), and ARIMA(0,1,1), respectively, for 10, 15, and 20 readings. ARIMA(0,1,1) without a constant gives a simple exponential smoothing. This corrects the challenges of the random walk model where the nonstationary series with noise and fluctuations are smoothed out. Hence, it takes averages of the last few observations to forecast rather than the most recent observation. Predictions give a constant value with no variations.

Using the ARIMA and even grey system models, the following predictions were obtained as shown in Figure 5.

Temperature values for GM(1,1) and ARIMA predictions were taken for 15, 20, 25, and 35 data points as the training and the rest as the test set. The ARIMA models were ARIMA(0,1,0), ARIMA(1,2,0), and ARIMA(1,1,0), respectively, for 15, 25, and 35. The predictions and original values are presented in Figures 5–7. It is observed that GM(1,1) pro-

duces data predictions of monotonic increase or decrease with a constant factor and produces varying predicted values that lie within the maximum and minimum original data readings. From Figure 5, the maximum pH value is 12.4 and the minimum pH value is 9.32. Similarly, in Figure 6, the maximum and minimum values read from the conductivity sensor are 195 and 125, respectively. Finally, in Figure 7, the maximum and minimum values recorded from the temperature sensor are 35.7 and 30.6, respectively. The predictions of ARIMA presented nonseasonal differencing with a constant term from the last value of original data reading, except for predictions of temperature where predictions were of the first-order autoregression. This is because the training data is insufficient to model the variations seen in the original data set. To test the performance of the prediction models used, the mean absolute percentage error (MAPE), mean absolute deviation (MAD), and the root mean square error (RMSE) performance metrics were used. The MAPE is percentage difference between the original values and the predicted, while the MAD is the distribution of the predicted values and it shows how close or spread out the values are from each other. The RMSE tells how far

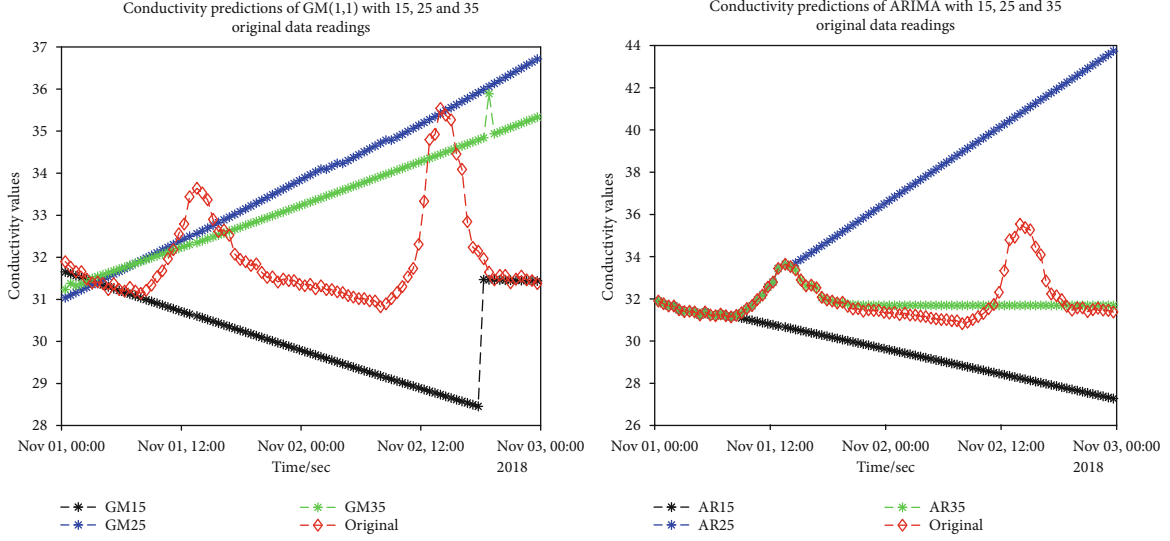


FIGURE 5: GM(1,1) and ARIMA predictions for varying data lengths in a 48-hour window for temperature readings.

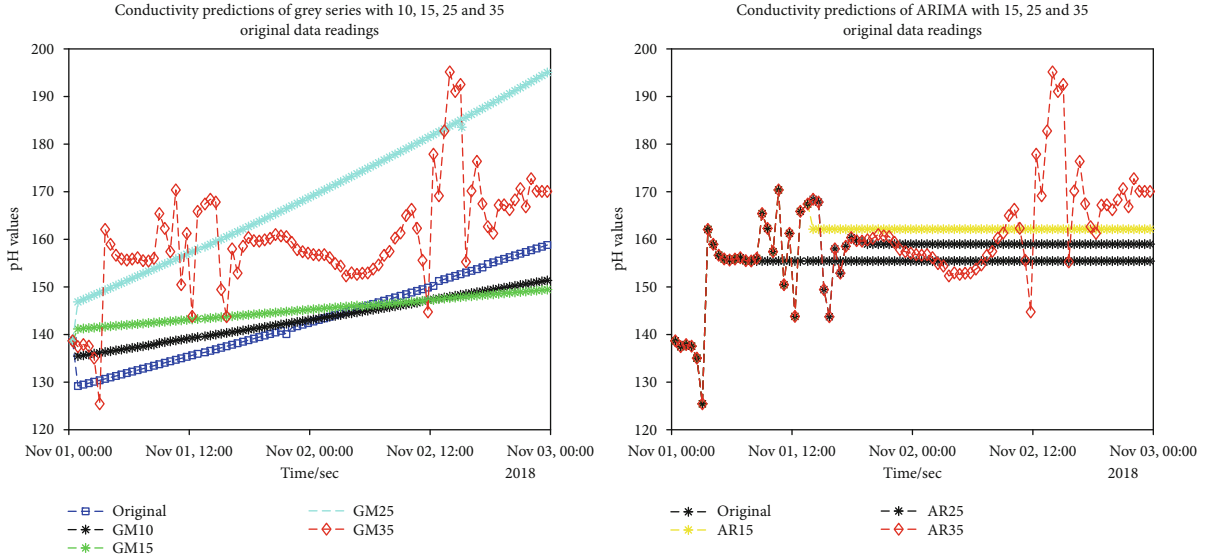


FIGURE 6: GM(1,1) and ARIMA predictions for varying data lengths in a 48-hour window of conductivity readings.

away original values are from the predicted values. The smaller RMSE value indicates the predicted model better fits the original data.

$$\begin{aligned}
 \text{MAPE} &= \left(\frac{1}{n}\right) * \sum \frac{|\text{actual} - \text{predicted}|}{|\text{actual}|} * 100, \\
 \text{MAD} &= \frac{1}{n} * \left(\sum |x_i - \bar{x}|\right), \\
 \text{RMSE} &= \sqrt{\left[\frac{\sum (\text{predicted} - \text{actual})^2}{n}\right]}.
 \end{aligned}
 \tag{14}$$

The MAPE values for pH with 15 original values for ARIMA were 2.08% more than those of GM, while those of temperature were 1.99% more for ARIMA and GM values,

respectively. For 25 original readings, ARIMA’s MAPE value was 2.85% more than GM. MAPE values are not the best metric for performance for low volume data. Where the MAPE is not a good indicator, the MAD or RMSE values are used. In this experiment, pH values and the RMSE values for 15 and 25 original values are 0.085 and 0.81 for ARIMA and 0.74 and 0.61 for GM. The RMSE for 15 values of temperature were 2.0 and 1.16 for ARIMA and GM, respectively. This confirms the assertion that GM(1,1) is generally a better model for predicting values of small data set, i.e., below 25 data sets and in the case of GM(1,1) as small as 4 data sets.

3.1. Improving on GM(1,1). Given that GM(1,1) showed better predictions with better performances than ARIMA for small data sets that have data with stochastic patterns, we use heuristic approaches to improve on the GM(1,1).

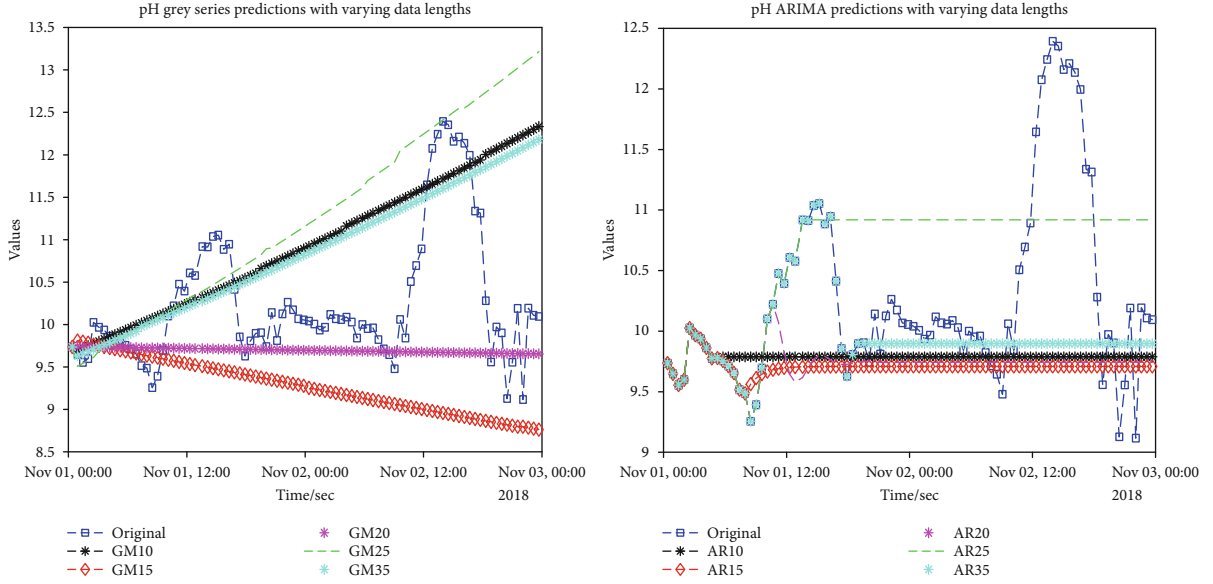


FIGURE 7: GM(1,1) and ARIMA predictions for varying data lengths in a 48-hour window of pH readings.

The reason is to find the optimum least training set that gives the best prediction over short intermittent periods. Two approaches are used in this paper, the heuristic and the modified rolling GM(1,1).

We use predictions for 20 data readings to estimate the optimum length of effective prediction.

The heuristic algorithm used is

If(predicted values < average(max, min)), predicted values,

$$\text{else, predicted values} + \frac{\sum_{i=1}^n x_i - m(X)\forall}{n}. \quad (15)$$

Graphs of original and adjusted values of pH, conductivity, and temperature are presented in Figure 8.

To obtain an optimum smallest data required for a good prediction and the optimum length of prediction that reveals the sudden changes that may occur in moving river, using data sets from pH, we forecast data of varying lengths of statistically minimum data sets (i.e., data sets below 30) and make forecasts to determine the optimum length of good forecasts.

3.2. The Rolling GM Approach. The grey prediction performs quantitative forecasts of data readings using unascertained characteristics of the data. This is done by using sequence operators of the original readings to generate and extract hidden patterns to establish a model for future predictions. When the data is chaotic and does not provide a good fit for most models, modifications are made to the original GM(1,1). One such heuristic approach is the rolling GM(1,1) [30]. The rolling grey series makes a forecast of time series data values using a constant window size of past data. A constant window size means it uses a fixed number of values for prediction. The rolling GM(1,1) is defined by the original GM(1,1) basic equation in the following equation.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b. \quad (16)$$

A constant window size is predefined by the user, such that for each round of prediction, for example, for original data sets

$$x^{(0)}(k), x^{(0)}(k+1), x^{(0)}(k+2), \dots, x^{(0)}(n). \quad (17)$$

The model predicts

$$x^{(0)}(k+1), x^{(0)}(k+2), x^{(0)}(k+3), \dots, x^{(0)}(n+1). \quad (18)$$

This means that for each round of prediction, k is increased by 1 $x^{(0)}(k+1), x^{(0)}(k+2), x^{(0)}(k+3), x^{(0)}(k+4)$.

And the preceding data predict the next round of values as follows:

$$x^{(0)}(k+2), x^{(0)}(k+3), x^{(0)}(k+4), x^{(0)}(k+5). \quad (19)$$

In [30], the equation of rolling GM(1,1) is defined in the following equation.

$$\hat{X}^{(0)}(r) = \left[x^{(0)}(r-n) - \frac{b}{a} \right] e^{-a} + \frac{b}{a}, \quad (20)$$

where n is the window size and r is the round whose value is to be predicted. Using a heuristic approach, we modify the constant window such that assuming a constant window size of n forecasting values in the i iteration

$$X^{(1)}(i) = 1 - e^a \left(x^{(0)}(kn) - \frac{b}{a} \right) e^{-ai}, \quad (21)$$

for $k = 0, 1, 2, \dots, n$, and k is increased by 1 when $i \% n = 0$, and $i > n$.

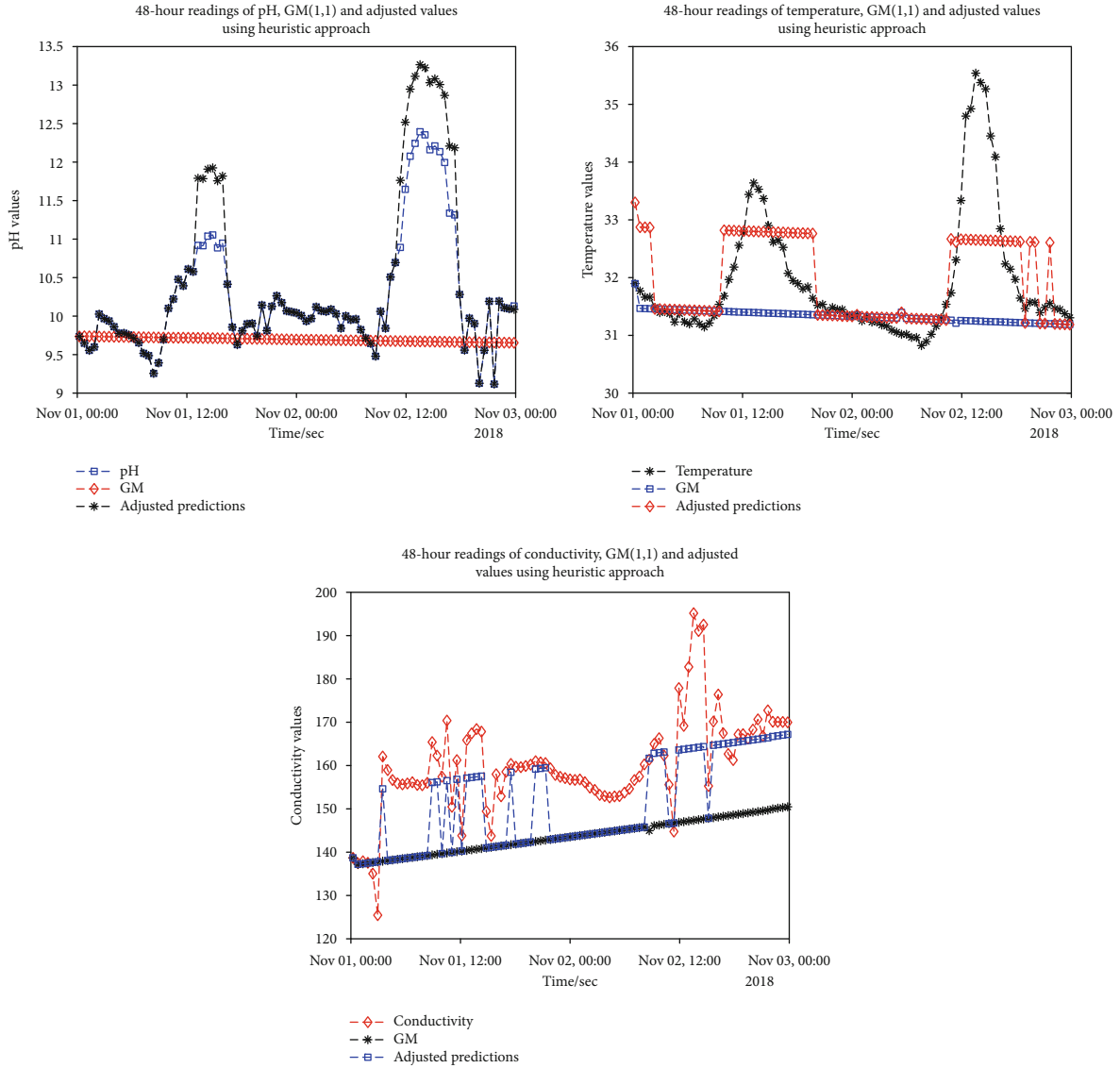


FIGURE 8: Graphs of pH, conductivity, and temperature using the heuristic approach to improving on GM(1,1).

In our heuristic approach, inverse AGO does not predict the original values used as the training set. Equation (21) becomes the basic equation for the proposed approach for prediction. The prediction after every round of kn selects the immediate data values from $kn + n$ as the training set for the next iteration. Predicted values do not include the period of the original data. This reduces the computational cost of forecasting the prediction period and original data as seen in the original grey series.

Using the minimum allowable data set for predicting in the GM(1,1) (i.e., 4 being the minimum), varying readings of between 4 and 10 training data readings are made for predictions of between 2 and 10 as presented in Figures 9-12.

Figure 9 shows a modified GM(1,1) approach using a rolling GM approach. The predicted values make substitutes of the original values intermittently to improve on the prediction accuracy of the model.

In Figure 10, different lengths of predictions are made for GM(1,1) using 6 original values alternatingly.

In Figure 11, different lengths of predictions are made for GM(1,1) using 8 original values alternatingly.

In Figure 12, different lengths of predictions are made for GM(1,1) using 10 original values alternatingly.

In Figure 13 and Table 1, we compare the relative errors of the values 4, 6, 8, and 10 when we forecast values of 2, 4, 6, 8, and 10. The best relative error values were found at 4 data points predicting 2 values, 6 data points predicting 2 and 4 values, and 8 data points predicting 2 and 6 values. Comparing with the corresponding energy savings per data point per predicting length, 4 data points predicting for 6 values give the best relative error vis-à-vis the energy savings. The energy savings were calculated based on the number of nodes transmitted versus the number of predictions made, such that no transmission is required during the period of prediction.

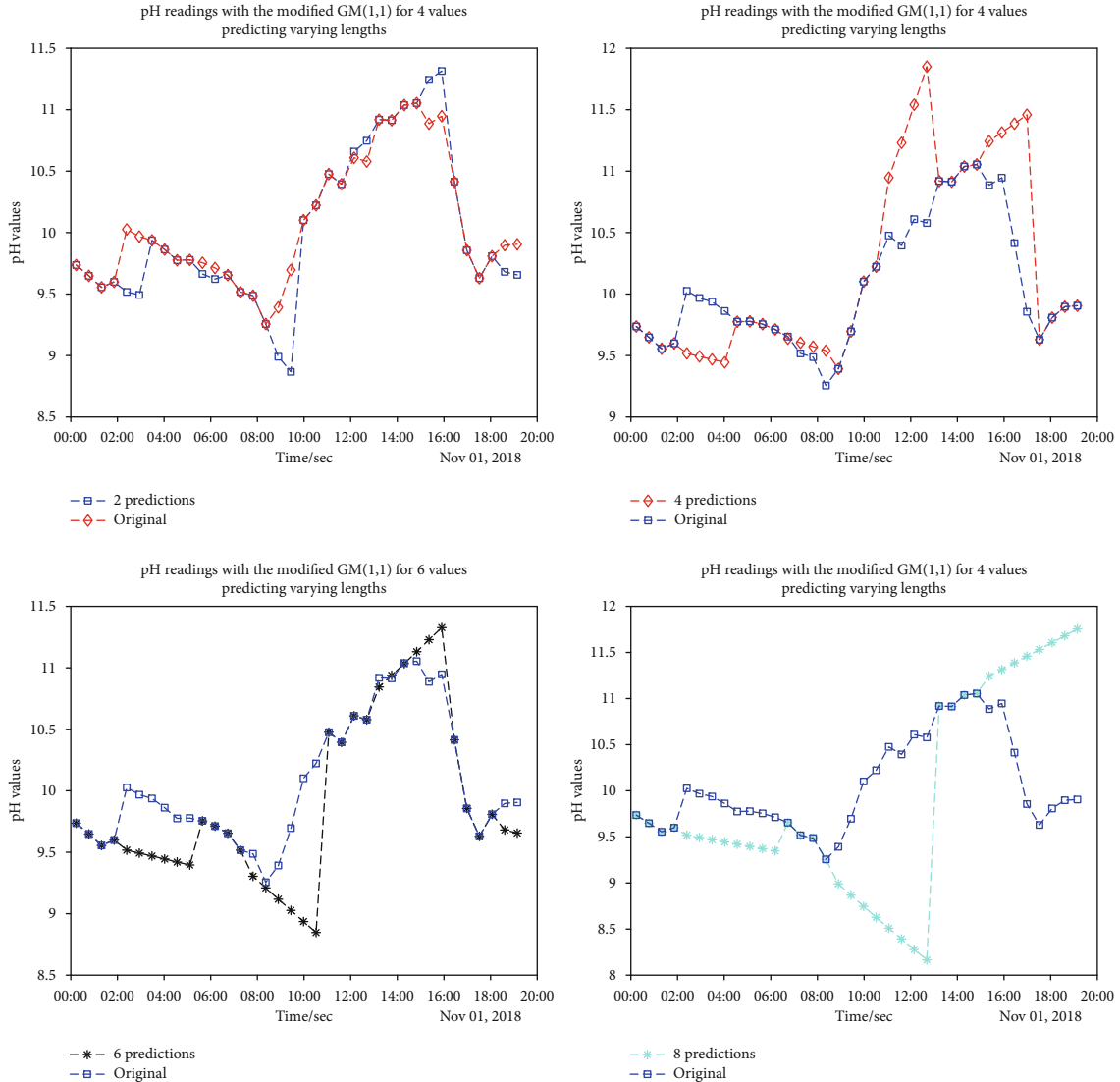


FIGURE 9: Predictions of 4 values forecasting different lengths.

For example, if 4 values are transmitted and 4 values are predicted, then the node saves 100% of its energy during the prediction period. The best relative errors were obtained when 4 values were transmitted with 6 values predicted, which is as follows:

$$\text{Energy savings} = \frac{\text{Original readings}}{\text{Predicted readings}} * 100\%. \quad (22)$$

This suggests that 150% of the energy of the nodes is saved for each original and predicted value cycle. This energy saving does not depend on other network conditions that affect energy loss but is only dependent on the ratio of energy losses due to transmission and reception.

4. Conclusions

Requiring large data sets for forecasting prediction models may not be ideal for WSN due to the cost of data transmis-

sion. GM(1,1) makes a good prediction with statistically small data sets, such as data set below 30. The minimum required for GM(1,1) is 4 which reduces the energy spent on transmission. A sensitivity analysis performed with GM(1,1) of varying lengths of data beginning at 4 to 10 training data set showed better performance of GM(1,1). Two approaches were used to improve the performance of GM(1,1) to provide near accurate predictions for shorter periods in a continuous stream of data. The two approaches used were the heuristic and the rolling GM. The heuristic approach is to reduce the length of prediction and improve the performance by modifying the inverse AGO approach of GM(1,1) to exclude the training data sets in prediction. The heuristic approach then performs a rolling GM(1,1) to continuously predict real-time stochastic data sets while maintaining the required reliability. Comparing the percentage energy savings vis-a-vis, the relative errors of the predictions and their lengths show the minimum of 4 data points is ideal if predictions are rolled for a constant window of 6

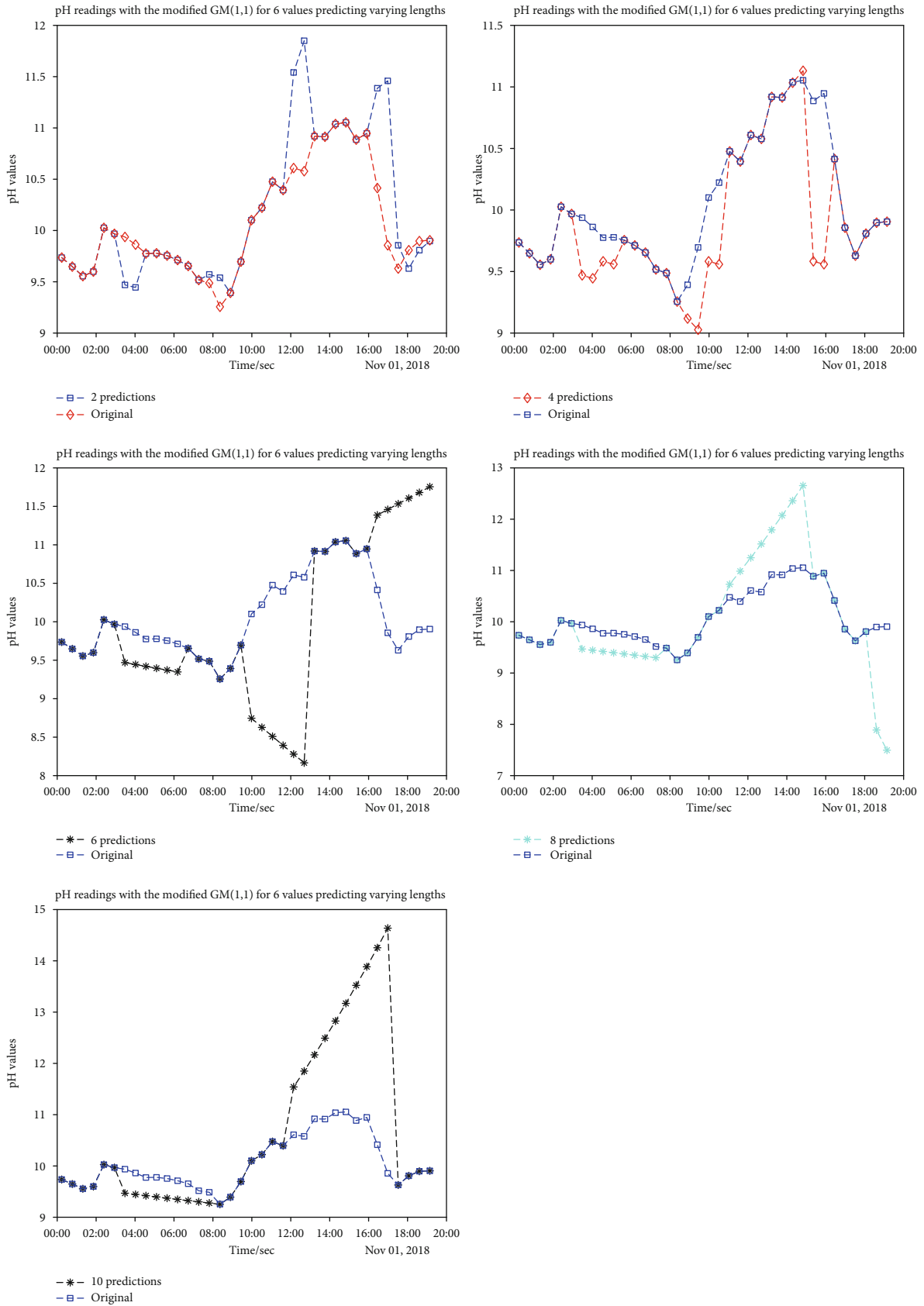


FIGURE 10: Predictions of 6 values forecasting different lengths.

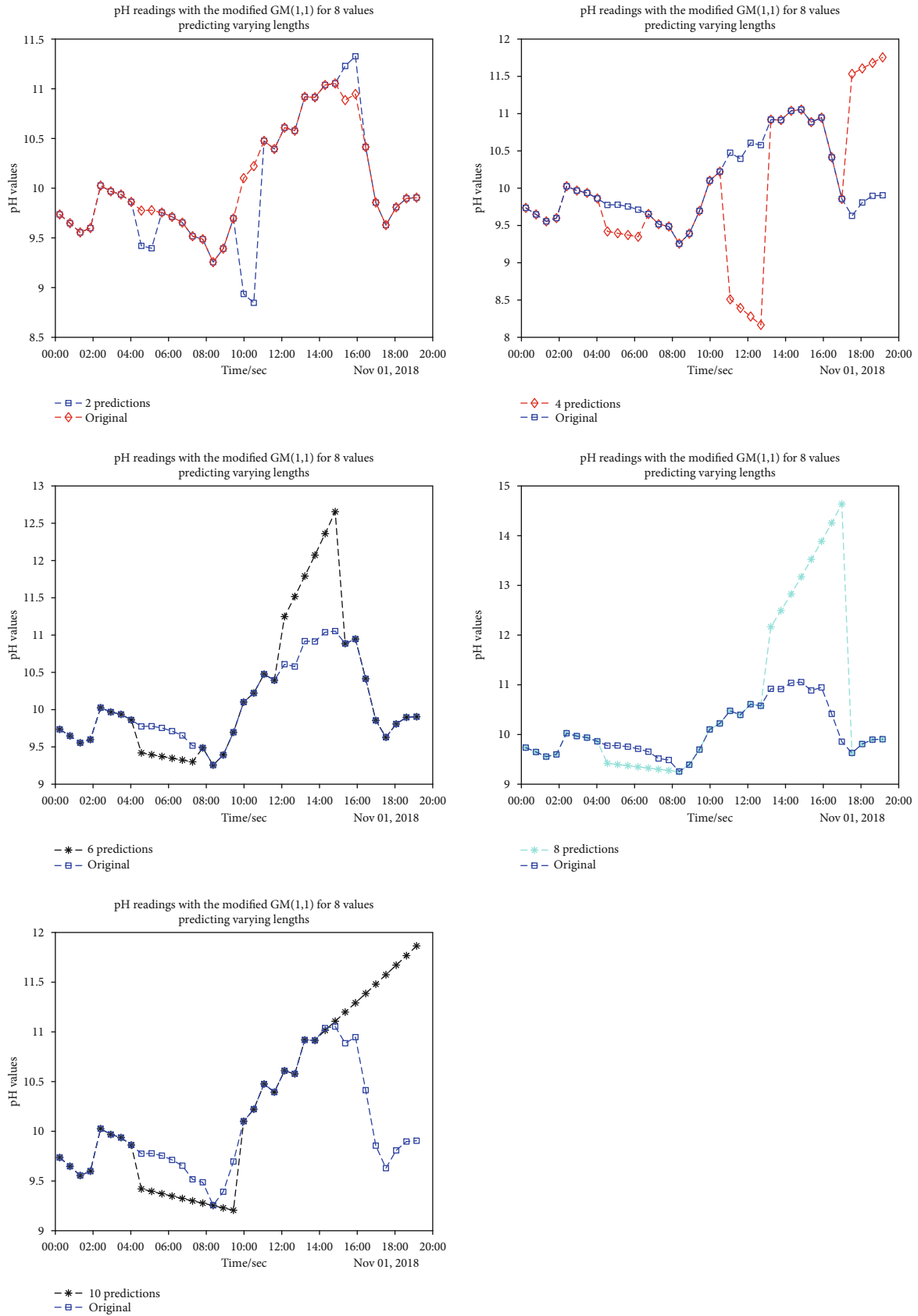


FIGURE 11: Predictions of 8 values forecasting different lengths.

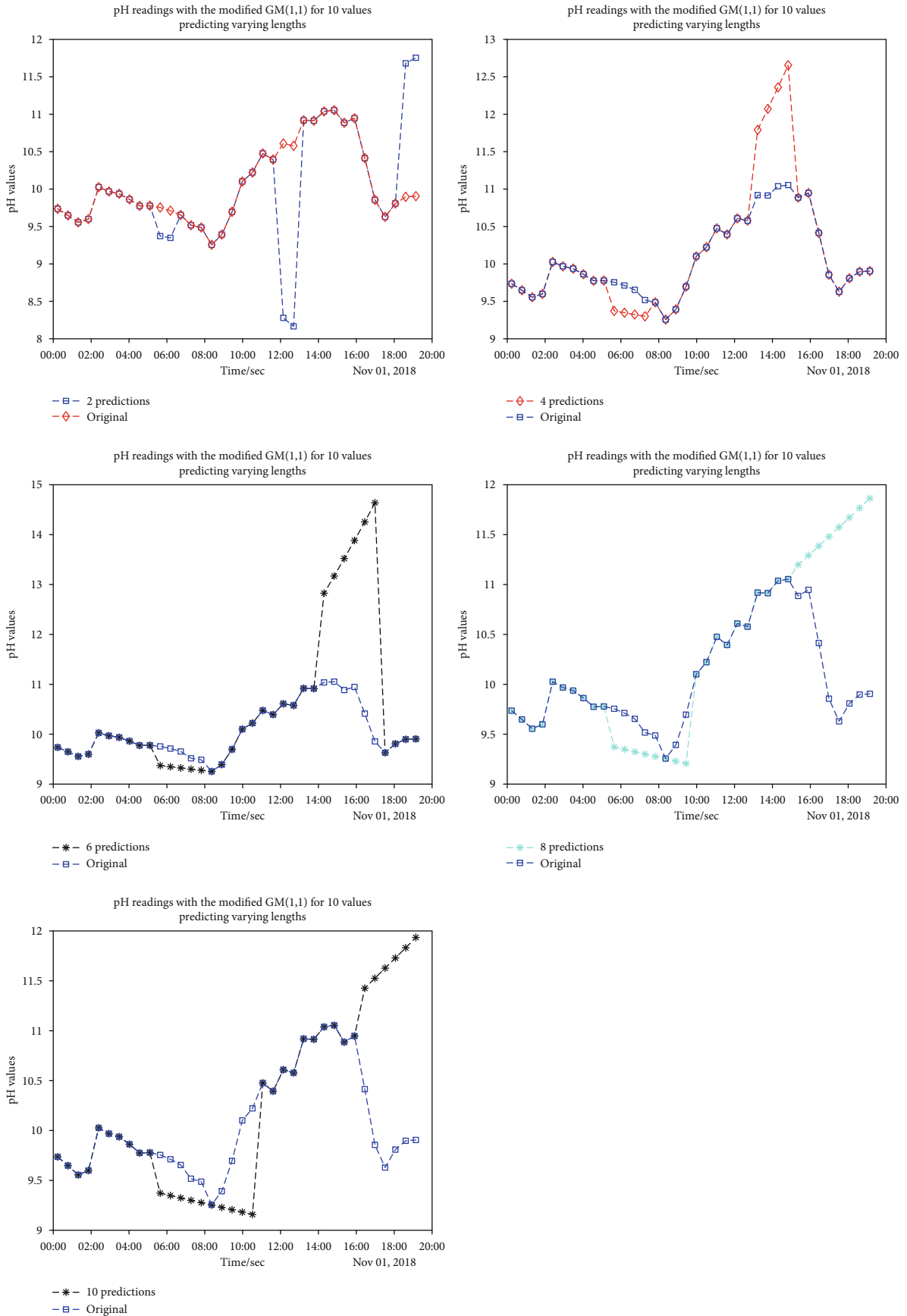


FIGURE 12: Predictions of 10 values forecasting different lengths.

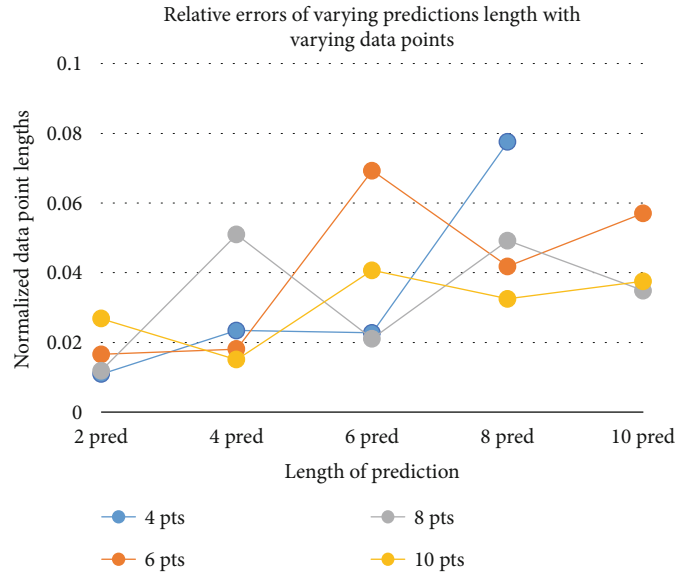


FIGURE 13: Graph of different lengths of training sets and different lengths of prediction.

TABLE 1: Energy savings per prediction.

Number of data points	Percentage of energy savings per transmission				
	2 pred	4 pred	6 pred	8 pred	10 pred
4	50	100	150	200	250
6	33.33	66.67	100	133.33	166.67
8	25	50	75	100	125
10	20	40	60	80	100

prediction lengths. This suggests a 150% savings in energy while ensuring the optimum accuracy of data predicted and the reliability of continuous data received at the receiving node. For future works, a network simulator or real-testbed experiment with real-time values may be modeled with these approaches to verify the proposed approach.

Data Availability

All necessary data is available in this transcript. The other data are available in the thesis submitted to the University of Ghana.

Conflicts of Interest

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

References

- [1] S. Srivastava, M. Singh, and S. Gupta, "Wireless sensor network: a survey," in *2018 International Conference on Automation and Computational Engineering (ICACE)*, pp. 159–163, Greater Noida, India, October 2018.
- [2] H. Toral-Cruz, F. Hidoussi, D. E. Boubiche, R. Barbosa, M. Voznak, and K. I. Lakhtaria, "A survey on wireless sensor networks," in *Next Generation Wireless Network Security and Privacy*, pp. 171–210, IGI Global, 2015.
- [3] H.-Y. Zhou, D.-Y. Luo, Y. Gao, and D.-C. Zuo, "Modeling of node energy consumption for wireless sensor networks," *Wireless Sensor Network*, vol. 3, no. 1, pp. 18–23, 2011.
- [4] K. S. Adu-Manu, C. Tapparello, W. Heinzelman, F. A. Katsriku, and J.-D. Abdulai, "Water quality monitoring using wireless sensor Networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 1, pp. 1–41, 2017.
- [5] F. Engmann, F. A. Katsriku, J.-D. Abdulai, K. S. Adu-Manu, and F. K. Banaseka, "Prolonging the lifetime of wireless sensor networks: a review of current techniques," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8035065, 23 pages, 2018.
- [6] K. S. Adu-Manu, N. Adam, C. Tapparello, H. Ayatollahi, and W. Heinzelman, "Energy-harvesting wireless sensor networks (EH-WSNs)," *ACM Transactions on Sensor Networks*, vol. 14, no. 2, pp. 1–50, 2018.
- [7] 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.
- [8] T. Rault, A. Bouabdallah, and Y. Challal, "Energy efficiency in wireless sensor networks: a top-down survey," *Computer Networks*, vol. 67, pp. 104–122, 2014.
- [9] A. Verma, M. P. Singh, J. P. Singh, and P. Kumar, "Survey of MAC protocol for wireless sensor networks," in *2015 Second International Conference on Advances in Computing and Communication Engineering*, pp. 92–97, Dehradun, India, May 2015.
- [10] R. C. Carrano, D. Passos, L. C. S. Magalhaes, and C. V. N. Albuquerque, "Survey and taxonomy of duty cycling mechanisms in wireless sensor networks," *IEEE Communication Surveys and Tutorials*, vol. 16, no. 1, pp. 181–194, 2014.
- [11] S. Kumar and S. Chauhan, "A survey on scheduling algorithms for wireless sensor networks," *International Journal of Computer Applications*, vol. 20, no. 5, pp. 7–13, 2011.
- [12] A. Abdelgawad and M. Bayoumi, "Data fusion in WSN," in *Resource-Aware Data Fusion Algorithms for Wireless Sensor*

- Networks*, Lecture Notes in Electrical Engineering, pp. 17–35, Springer, Boston, MA, USA, 2012.
- [13] G. Dhand and S. S. Tyagi, “Data aggregation techniques in WSN: survey,” *Procedia Computer Science*, vol. 92, pp. 378–384, 2016.
- [14] M. Huang, K. Zhang, Z. Zeng, T. Wang, and Y. Liu, “An AUV-assisted data gathering scheme based on clustering and matrix completion for smart ocean,” *IEEE Internet of Things Journal*, 2020.
- [15] B. Jiang, G. Huang, T. Wang, J. Gui, and X. Zhu, “Trust based energy efficient data collection with unmanned aerial vehicle in edge network,” *Transactions on Emerging Telecommunications Technologies*, no. article e3942, 2020.
- [16] F. Zhou, Z. Chen, S. Guo, and J. Li, “Maximizing lifetime of data-gathering trees with different aggregation modes in WSNs,” *IEEE Sensors Journal*, vol. 16, no. 22, pp. 8167–8177, 2016.
- [17] U. Raza, A. Camerra, A. L. Murphy, T. Palpanas, and G. P. Picco, “Practical data prediction for real-world wireless sensor networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2231–2244, 2015.
- [18] K. S. Adu-Manu, F. Katsriku, J.-D. Abdulai, J. Marx Gómez, and W. Heinzelmann, “Network lifetime maximization with adjustable node transmission range,” in *Smart Cities/Smart Regions – Technische, wirtschaftliche und gesellschaftliche Innovationen*, pp. 693–707, Springer Vieweg, Wiesbaden, Germany, 2019.
- [19] J. Roselin, P. Latha, and S. Benitta, “Maximizing the wireless sensor networks lifetime through energy efficient connected coverage,” *Ad Hoc Networks*, vol. 62, pp. 1–10, 2017.
- [20] A. Förster, “Machine learning techniques applied to wireless ad-hoc networks: guide and survey,” in *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp. 365–370, Melbourne, Qld, Australia, December 2007.
- [21] D. Praveen Kumar, T. Amgoth, and C. S. R. Annavarapu, “Machine learning algorithms for wireless sensor networks: a survey,” *Information Fusion*, vol. 49, pp. 1–25, 2019.
- [22] H. He, Z. Zhu, and E. Makinen, “A neural network model to minimize the connected dominating set for self-configuration of wireless sensor networks,” *IEEE Transactions on Neural Networks*, vol. 20, no. 6, pp. 973–982, 2009.
- [23] Q. Yu, L. Jibin, and L. Jiang, “An improved ARIMA-based traffic anomaly detection algorithm for wireless sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, Article ID 9653230, 2016.
- [24] T. H. D. Ngo and W. Bros, “The Box-Jenkins methodology for time series models,” in *Proceedings of the SAS Global Forum 2013 conference*, pp. 1–11, San Francisco, 2013.
- [25] E. Kayacan, B. Ulutas, and O. Kaynak, “Grey system theory-based models in time series prediction,” *Expert Systems with Applications*, vol. 37, no. 2, pp. 1784–1789, 2010.
- [26] S. Diwakaran, B. Perumal, and K. Vimala Devi, “A cluster prediction model-based data collection for energy efficient wireless sensor network,” *The Journal of Supercomputing*, vol. 75, no. 6, pp. 3302–3316, 2019.
- [27] S. Kashyap, E. Björnson, and E. G. Larsson, “On the feasibility of wireless energy transfer using massive antenna arrays,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3466–3480, 2016.
- [28] P. Ganjewar, S. Barani, and S. J. Wagh, “A hierarchical fractional LMS prediction method for data reduction in a wireless sensor network,” *Ad Hoc Networks*, vol. 87, pp. 113–127, 2019.
- [29] X. Luo and X. Chang, “A novel data fusion scheme using grey model and extreme learning machine in wireless sensor networks,” *International Journal of Control, Automation and Systems*, vol. 13, no. 3, pp. 539–546, 2015.
- [30] D. P. Singh, V. Bhateja, and S. K. Soni, “Energy optimization in WSNs employing rolling grey model,” in *2014 International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 801–808, Noida, India, February 2014.
- [31] S. K. Soni, N. Chand, and D. P. Singh, “Reducing the data transmission in WSNs using time series prediction model,” in *2012 IEEE International Conference on Signal Processing, Computing and Control*, pp. 1–5, Wanknaghat Solan, India, March 2012.
- [32] I. Ben Arbi, F. Derbel, and F. Strakosch, “Forecasting methods to reduce energy consumption in WSN,” in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, Turin, Italy, May 2017.
- [33] V. Katiyar, N. Chand, and S. Soni, “Grey system theory-based energy map construction for wireless sensor networks,” in *International Conference on Advances in Computing and Communications*, pp. 122–131, Kochi, India, 2011.
- [34] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, 2003.
- [35] J. Deng, “Introduction to grey system,” *Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.
- [36] S. Liu, Y. Yang, N. Xie, and J. Forrest, “New progress of grey system theory in the new millennium,” *Grey Systems: Theory and Application*, vol. 6, no. 1, pp. 2–31, 2016.
- [37] A. K. Singh, N. Purohit, and S. Varma, “Fuzzy logic based clustering in wireless sensor networks: a survey,” *International Journal of Electronics*, vol. 100, no. 1, pp. 126–141, 2013.
- [38] R. Arthi and K. Murugan, “Localization in wireless sensor networks by hidden Markov model,” in *ICoAC 2010*, pp. 14–18, Chennai, India, December 2010.
- [39] G. Tan, J. Yan, C. Gao, and S. Yang, “Prediction of water quality time series data based on least squares support vector machine,” in *Procedia Engineering*, vol. 31, pp. 1194–1199, 2012.
- [40] A. Shareef, Y. Zhu, and M. Musavi, “Localization using neural networks in wireless sensor networks,” in *Proceedings of the 1st International ICST Conference on Mobile Wireless Middleware, Operating Systems and Applications*, pp. 1–7, Innsbruck Austria, 2009.

Research Article

The Hybrid Solar-RF Energy for Base Transceiver Stations

Cuong V. Nguyen,¹ Minh T. Nguyen ,² Toan V. Quyen,³ Anh M. Le,⁴ and Linh H. Truong⁵

¹*Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam*

²*Department of Electrical Engineering, Thai Nguyen University of Technology, Thai Nguyen 24000, Vietnam*

³*School of Electronics Engineering, Kyungpook National University, Daegu 41566, Republic of Korea*

⁴*College of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu 30013, Taiwan*

⁵*Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 30013, Taiwan*

Correspondence should be addressed to Minh T. Nguyen; tuanminh.nguyen@okstate.edu

Received 15 April 2020; Revised 11 June 2020; Accepted 17 June 2020; Published 14 July 2020

Academic Editor: Yujin Lim

Copyright © 2020 Cuong V. Nguyen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The base transceiver stations (BTS) are telecom infrastructures that facilitate wireless communication between the subscriber device and the telecom operator networks. They are deployed in suitable places having a lot of freely propagating ambient radio frequency (RF) and solar energies. This paper is aimed at converting received ambient environmental energy into usable electricity to power the stations. We proposed a hybrid energy harvesting system that can collect energy from RF and solar energies at the same time. The sources are combined to provide to a significant amount, to contribute to operational expenditures that reduce energy costs, and to improve the energy efficiency of the base station sites in rural areas from the most common renewable resources since the base stations are major consumers of cellular networks. The hybrid systems are designed with circuits, simulated, and compared to show their good performance to the base stations. PSIM, PROTEUS, and MATLAB software are used to simulate for evaluating the voltage and the current output of the hybrid systems that meet the power requirements. The design and simulation results show the feasibility of our proposed method with the battery storage that can be deployed not only in real base stations but also for other electrical operated systems.

1. Introduction

The wireless communication system is one of the most important technologies for promoting economic and social development around the globe. Cellular systems, such as long-term evolution (LTE), are designed assuming that mobile devices connect to a base station to communicate. The base stations receive and transmit data from and to mobile users, called base transceiver stations (BTS). Since the telecom infrastructures are increasing everywhere, there are many base stations being deployed to satisfy the wireless service coverage. In particular, more than five billion new customers using wireless mobile networking services across the world were added by 2010 [1]. And according to the

forum of Wireless World Research, in 2020, seven trillion wireless devices will serve seven billion people [2]. However, the extensive growth in the number of users, the new wireless products, the demand of the quality of services, and the rising service usage times result in the increased energy consumption for such networks, especially for the BTS. There are many research results considering energy-efficient models for both operation and embodied energy consumption for cellular networks [3–5]. However, the world is still expecting to have further energy-efficient approaches for such networks.

Recently, the BTS are deployed in many places including rural and remote areas that may not have sustainable grid electricity systems. This motivates researchers seeking for

TABLE 1: General levels of required power at some base transceiver stations.

Types of BTS	Demand of power (kW)
GSM base station 2/2/2	1.8–2.0
GSM base station 4/4/4	2.3–3.5
UMTS node B macro/fiber 4/4/4 base station	1.7–2.0
Large WiMAX base station	1.6–1.8

other resources to meet the power requirements. In general, the required power to support BTS is addressed in Table 1. In addition, the more the number of wireless communication services is, the more power would be consumed at the stations in the networks.

There are at least two strong points to motivate using green or renewable energy resources. First, this is vital to minimize the environmental impacts on climate change, caused by CO₂ and other greenhouse gases in the atmosphere, emitted from the use of fossil fuels as primary resources to produce electrical energy. All network providers need to reduce electricity bills for competitive purposes. Second, when the use of energy harvesting technologies is increasing, the dependence on the electricity grid or fossil fuels will decrease significantly. This has led to a reduction in operating costs for stations aimed at reducing service costs. So, energy harvesting technologies are considered to support the networks, not only at the stations but also at other low-energy consumption devices.

Energy harvesting (EH) or energy scavenging is a very promising technology for applications where batteries are impractical and/or power grid networks are not sustainable. The energy is achieved from ambient sources, such as solar, vibration, thermal, wind, and radio frequency (RF). The amount of harvested energy strongly depends on the performance, specific materials, and ambient resources as shown in Table 2 [6]. Based on that, many applications have applied different EH technologies to improve their performance and service quality. Currently, the harvested energy can support many electricity-operated systems, such as wireless sensor networks [6, 7], low-power consumption devices, robots, unmanned aerial vehicles [8–10], and even wheelchairs [11] to prolong the operating time. EH technologies can be combined to collect energy not only from a single source but also from some suitable ambient sources, called hybrid, to provide sufficient power for specific applications [12, 13].

In this paper, we propose a hybrid EH system to power the BTS as a backup system after the batteries and the generators may not perform well. As shown in Figure 1, the hybrid system can work together with other backup systems to provide the best solution for powering the BTS. The two most common resources such as RF and solar energies are collected to support the system. Different from the existing work, the proposed method provides a sufficient amount of energy to support the working BTS as well as the sustainable grid in a good condition. We design the systems in different

TABLE 2: Energy scavenging from different resources [6].

Energy sources	Conditions	Performance
Solar	Outdoor	7500 $\mu\text{W}/\text{cm}^2$
Solar	Indoor	100 $\mu\text{W}/\text{cm}^2$
RF	WiFi	0.001 $\mu\text{W}/\text{cm}^2$
RF	GSM	0.1 $\mu\text{W}/\text{cm}^2$
Vibration	1 m/s ²	100 $\mu\text{W}/\text{cm}^3$
Thermal	$\Delta T = 5^\circ\text{C}$	60 $\mu\text{W}/\text{cm}^2$

cases of using the stand-alone EH and hybrid EH and compare the effectiveness of the systems. All the circuits are designed and run to provide promising results. The outcomes are shown to be sufficient to power the BTS that supports the communication networks working well in good conditions.

The rest of the paper is addressed as follows. Section 2 provides the system model. Section 3 presents and analyzes the components of the hybrid solar-RF power supply system. The simulation and results are mentioned in Section 4. Finally, conclusions and future work are addressed in Section 5.

2. System Model

In this section, the full picture of the hybrid system is modeled with all components for the purpose of harvesting energy to support the BTS. Figure 2 shows the model system of hybrid power. The hybrid system consists of main components such as the solar cells, the RF harvester, a common DC bus, the stabilizer system, and the backup batteries. The solar energy system is composed of the photovoltaic (PV) arrays. They apply the maximum power point tracking (MPPT) method to adjust the duty cycle of the boost converter to achieve high voltage.

The RF energy is received by the antenna from the ambient environment. The main function of the matching circuit is to step up the input voltage of the rectifier and diminish the transmission loss from the antenna to the rectifier. The rectifier is used to convert AC-DC voltage. After that, the voltage outputs of the solar cells and RF energy will be supplied to the boost converter to step up these voltages to a higher level. The stabilizer system will receive the output voltage of the boost converter and produce a suitable voltage for the BTS system. The backup battery system is used to store the power when the hybrid system is at the peak power point and then used when the power system is not available.

3. The Proposed Hybrid System

In this section, all the main components are addressed in detail with the related functions and equations.

3.1. The DC-DC Boost Converter. The DC-DC boost converter is used to step up the input voltage to the higher voltage [14]. The boost converter configuration includes the primary input power, an inductor, the MOSFET transistor,

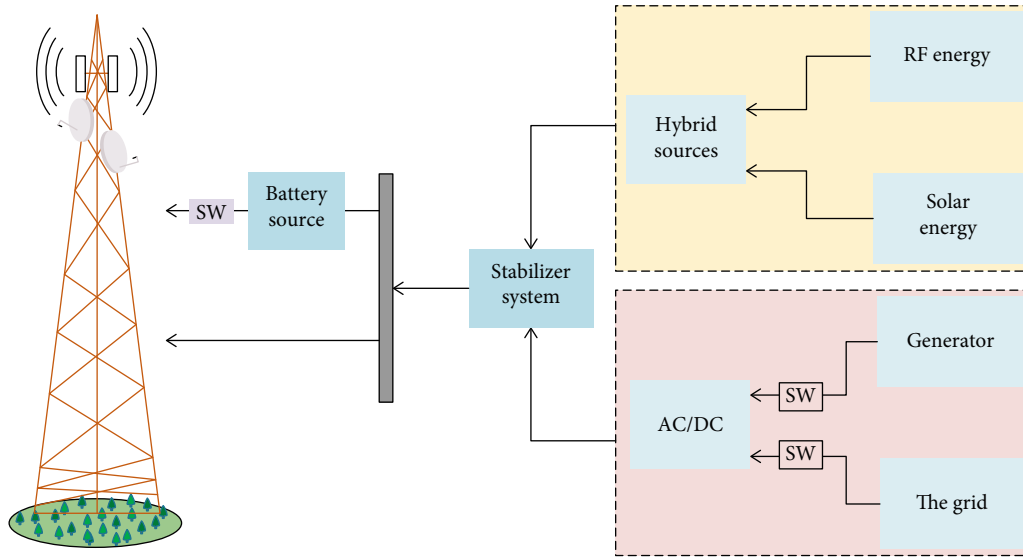


FIGURE 1: Hybrid energy harvesting system model with solar and RF energy.

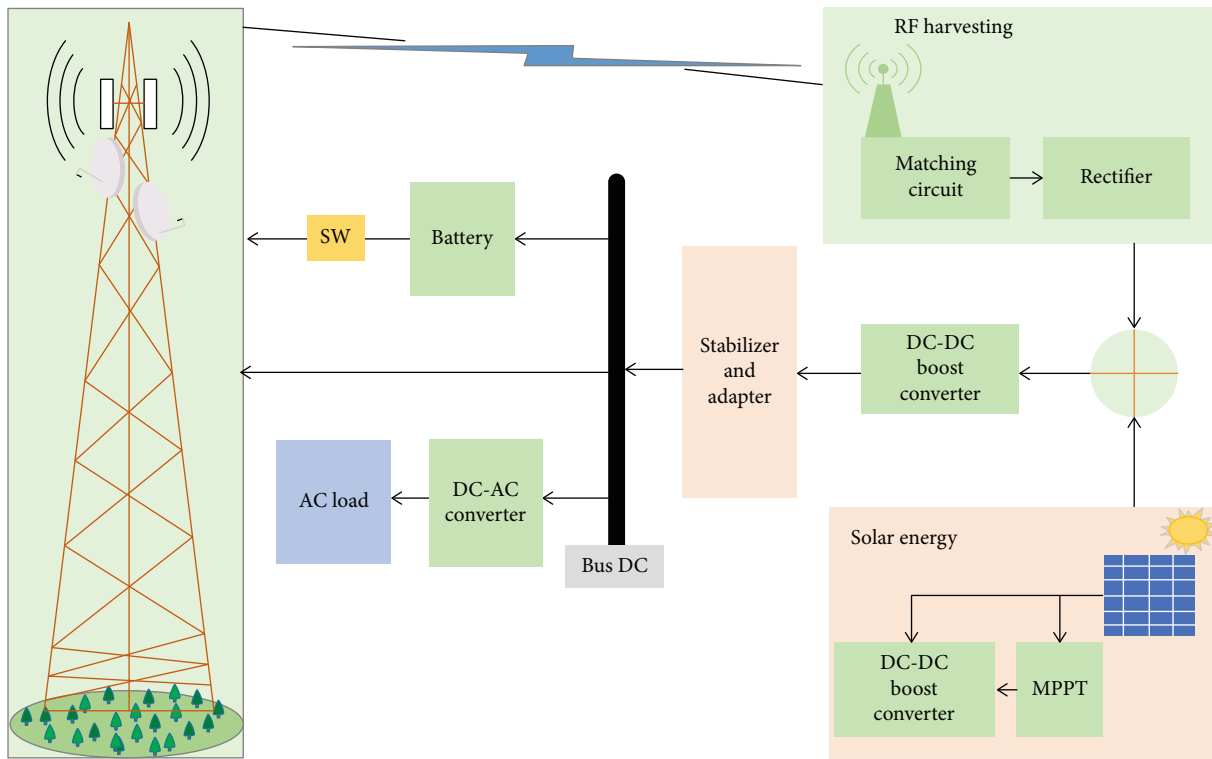


FIGURE 2: The hybrid system model with all components.

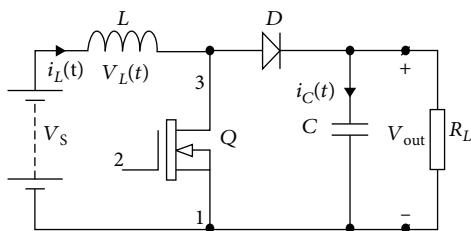


FIGURE 3: The scheme of the boost converter.

the capacitor, the diode, and the load, as shown in Figure 3. The MOSFET transistor turns on and turns off at a certain frequency and certain duty cycle. During MOSFET ON, the inductor will store the current flowing through it, and the diode prevents unnecessary discharge of the capacitor to the source. During MOSFET OFF, the capacitor will be fully charged by the source and provides power to the load later. The boost converter has two modes which are MOSFET transistor on and MOSFET transistor off. Here, V_s is the source

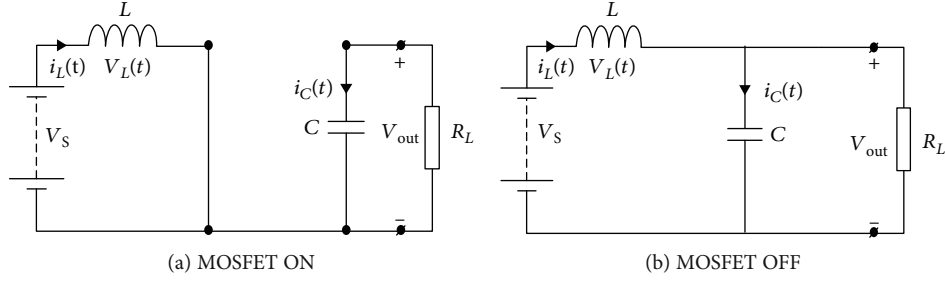


FIGURE 4: The MOSFET mode of the boost converter.

voltage of the boost converter, V_{out} is the output voltage, D_{on} is the on-percentage of the time, L is the inductance of the inductor, and D_{off} is the off-duty cycle. The t_{on} and t_{off} are the time interval of MOSFET ON or OFF, respectively.

The relationship between input and output voltage is as follows:

$$\frac{V_{out}}{V_s} = \frac{1}{D_{off}}. \quad (1)$$

Equally, the off-duty cycle may be expressed as follows:

$$D_{off} = 1 - D_{on}. \quad (2)$$

The effect of the inductor and the duty cycle is shown in the following formula:

$$\frac{V_s}{L} t_{on} = \frac{V_{out} - V_s}{L} t_{off}, \quad (3)$$

The MOSFET is like a conductor (the MOSFET transistor turns on), as shown in Figure 4(a). The input current is only going to flow to the inductor. The whole source power is stored in this inductor as indicated in Equation (5). During this period, the voltage is constant, but the current increases. Hence, the storing power of the inductor is gradually increasing.

$$V_L = V_s, \quad (4)$$

when the MOSFET transistor turns off, as shown in Figure 4(b). The supplying power is not only the input power but also the storing power of the inductor during MOSFET ON mode. Based on the energy conversion theory, when the supplying power rises, the power to provide the load will also increase as indicated in Equation (6). The supplying power will charge for the capacitor and supply to the load at the same time.

$$V_{out} = V_s + V_L. \quad (5)$$

3.2. The Buck Converter. Figure 5 depicts the scheme of the buck converter. The fundamental difference between the boost converter and the buck converter is the position variation of the three main electrical components such as the MOSFET transistor, inductor, and diode. This changes

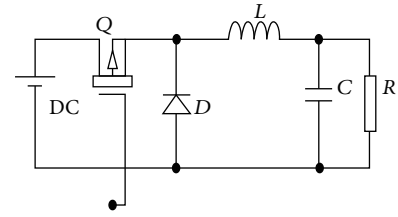


FIGURE 5: The scheme of the buck converter.

the feature of this circuit that converts the high voltage into low voltage [15]. The MOSFET transistor turns on and turns off at a certain frequency and certain duty cycle. An inductor L limits the flowing current to prevent the capacitor, but when the MOSFET changes from off to on, the current through the inductor cannot change instantaneously. The inductor will force the current to keep flowing through the switch even after we open it, which is an extremely dangerous phenomenon, so the diodes are used to keep the current flowing in the different path when the MOSFET is off. Last, the capacitor is used to store the voltage. The buck converter has two modes which are MOSFET transistor on and MOSFET transistor off. Here, V_s is the source voltage of the boost converter, V_{out} is the output voltage, D_{on} is the on-percentage of the time, L is the inductance of the inductor, and D_{off} is the off-duty cycle.

The effect of the inductor and the duty cycle is shown in the following formula:

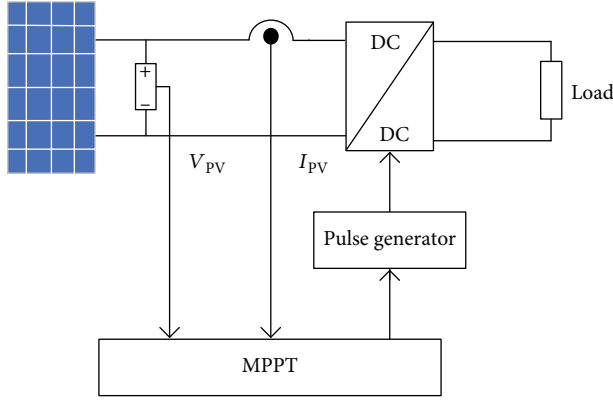
$$\frac{V_s - V_{out}}{L} t_{on} = \frac{V_{out}}{L} t_{off}, \quad (6)$$

$$\frac{V_{out}}{V_s} = \frac{t_{on}}{t_{on} + t_{off}} = \frac{t_{on}}{T_s} = D_{on}. \quad (7)$$

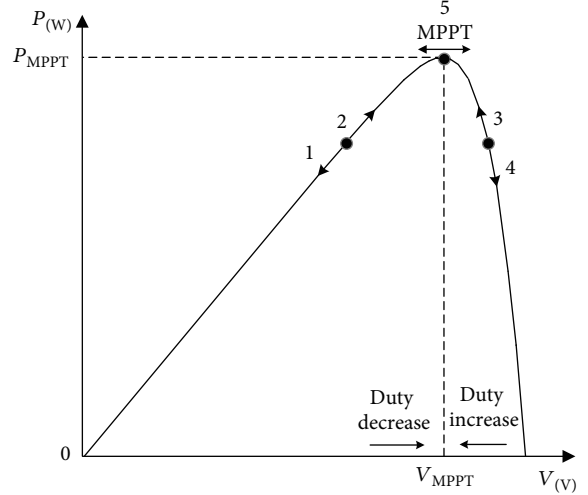
When the switch is off, the input current does not flow anywhere. The capacitor's current will flow through the load, the diode, and the inductor. This current keeps flowing through the inductor when the switch is off which helps prevent the dangerous phenomena from the instantaneously increasing current.

The off-duty cycle is expressed as

$$D_{off} = 1 - D_{on}. \quad (8)$$



(a) The block diagram



(b) The nonlinear characteristic

FIGURE 6: The perturb and observe (P&O) algorithms in the MPPT controller.

TABLE 3: The principle of operation for the P&O algorithms in MPPT.

Status	Condition	Actions
1	$\Delta P_{PV} < 0, \Delta V_{PV} < 0$	Duty decrease
2	$\Delta P_{PV} > 0, \Delta V_{PV} > 0$	Duty decrease
3	$\Delta P_{PV} > 0, \Delta V_{PV} < 0$	Duty increase
4	$\Delta P_{PV} < 0, \Delta V_{PV} > 0$	Duty decrease
5	$\Delta P_{PV} = 0, \Delta V_{PV} = 0$	No action

3.3. MPPT Algorithms. The basic and adaptive MPPT algorithms have been studied for distributed photovoltaic systems to maximize the energy production of a photovoltaic module. MPPT using a fixed step size, such as perturb and observe (P&O) and incremental conductance, suffer from a trade-off between tracking accuracy and tracking speed. In this paper, we use P&O algorithms. P&O algorithms operate on the fundamental principle that the variation (ΔP_{PV}) of PV voltage (V_{PV}) and the variation (ΔP_{PV}) of PV power (P_{PV}) become zero at the maximum power point (MPP). The principle of MPPT algorithms is shown in Figure 6 and Table 3.

3.4. The Voltage Multiplier. The voltage multiplier (VM) converts the AC input source from a lower voltage to a higher DC voltage. The VM includes diodes and a capacitor. Because the input is an AC source, the operation is divided into two loops, as shown in Figure 7.

In the first loop, the current flows through the diode D_1 and charges for the capacitor C_1 , but the current cannot charge for the diode C_2 because the current D_1 is blocked by the diode. After that, to the second loop, the current flows to the C_1 to the diode C_2 and charges for the diode C_2 , but the current is blocked by the diode C_1 . The output voltage equals the voltage of the diode C_2 , as shown in Equation (8), because the diode C_1 is discharged by the direction of the

current. The higher the numbers of the stages have, the higher the output voltage can achieve [16].

4. Analysis and Discussions

In this work, the proposed power system supplies power 48 V and 52 A to the BTS system. Because the settling time is inversely proportional to the output values, we need to trade off between the settling time and the output values. Our target harvested 20 VDC RF energy after using the voltage multiplier and 100 VDC solar energy after using the boost converter within the proper settling time. The current of the hybrid system is less than 2 A, but the desired current is much more than this result. Therefore, the output values of the hybrid system need to step up to great values in order to achieve a high output current. To create the 52 A and 48 VDC output values, the output values of the hybrid solar-RF energy are processed by the boost converter and stabilizer system. The DC-DC boost converter is used to step up the voltage to 400 VDC. The stabilizer and adapter system makes the desired current and voltage output stable. The problems are simulated and solved in PSIM, PROTEUS, and MATLAB Simulink.

4.1. The Hybrid Solar-RF Energy Harvesting. The solar and RF energy is abundant in the surrounding environment at the base transceiver station (BTS) system. Hence, the hybrid renewable energy harvesting includes solar energy and RF energy to leverage these plentiful sources. The hybrid system is able to harvest 20 VDC from RF energy and 100 VDC from solar cells.

4.2. RF Energy Harvesting. The voltage multipliers (VM) are used to convert AC to DC voltage and step up to higher voltage. The stage numbers of the VM directly affect the output voltage. The higher the stage numbers of the VM have, the higher the output voltage achieves [8], but the VM drawback is the delay time. In the RF system, we choose

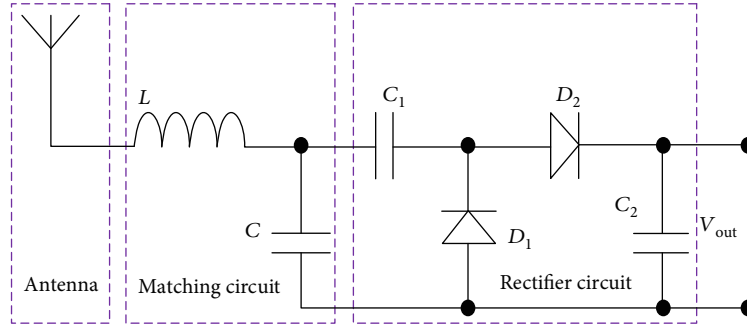


FIGURE 7: The scheme of the RF energy harvesting.

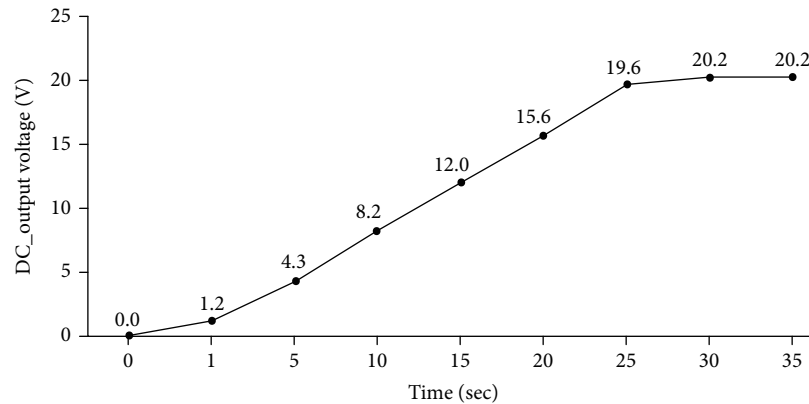


FIGURE 8: The output voltage of the RF energy system.

the 7-stage voltage multipliers to boost the harvesting voltage to the desired values. With the 7-stage VM, the longer the time the voltage multipliers take, the higher the output voltage is able to achieve [17]. In order to balance between the output voltage and the delay time, our target gets 20 DC voltage from the RF energy system. In Figure 8, the output voltage reaches 20 VDC in just 25 seconds, so the results are satisfying for both the desired output value and the delay time. Hence, the RF energy meets 20 DC voltage, approximately 20% of the energy of the hybrid system.

4.3. Solar Energy Harvesting. We assume that the voltage of the solar panel harvested from the ambient environment is 20 VDC. The harvesting voltage needs to be boosted to the higher voltage by the solar system. The solar system comprises the boost converter and maximum power point tracking (MPPT) to step up the voltage to the desired values with high efficiency [18–20]. Figure 9 presents the output values of the stand-alone solar PV system. The output voltage approaches 100 volts and the output current approaches 1.2 A for about 6 seconds, and those values are maintained until the end of the simulation. Approximately 80% of the energy of the hybrid system is powered by the solar PV system.

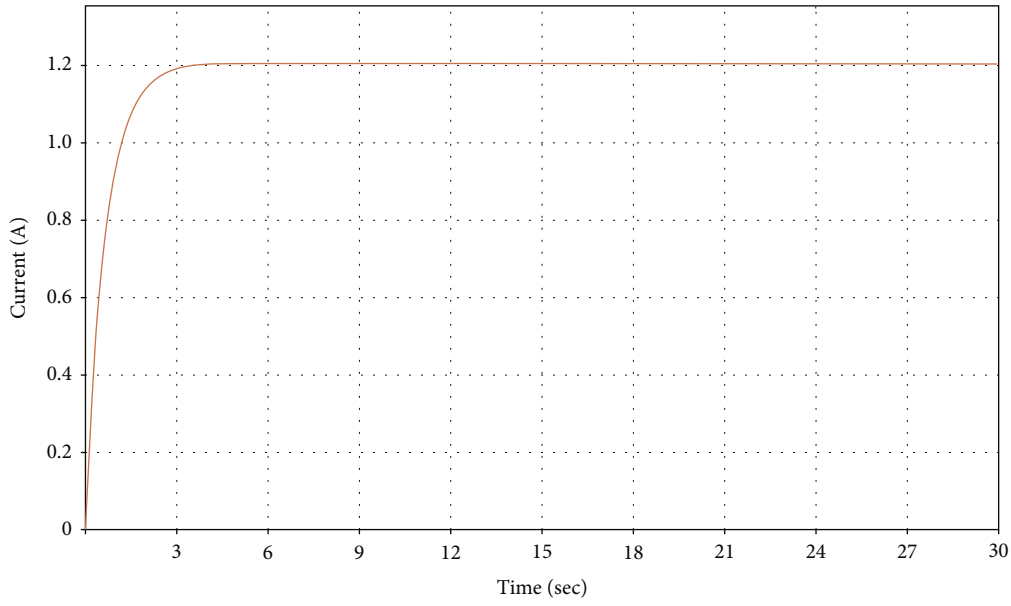
4.4. Boost Converter. The original voltage and current harvested from the ambient resources are not suitable to supply

for BTS systems that require high current and low DC voltage. In order to have a high output current, the input voltage and current will be stepped up to great values. By using the boost converter, the harvested power of 120 VDC will be boosted to 400 VDC and 82 A.

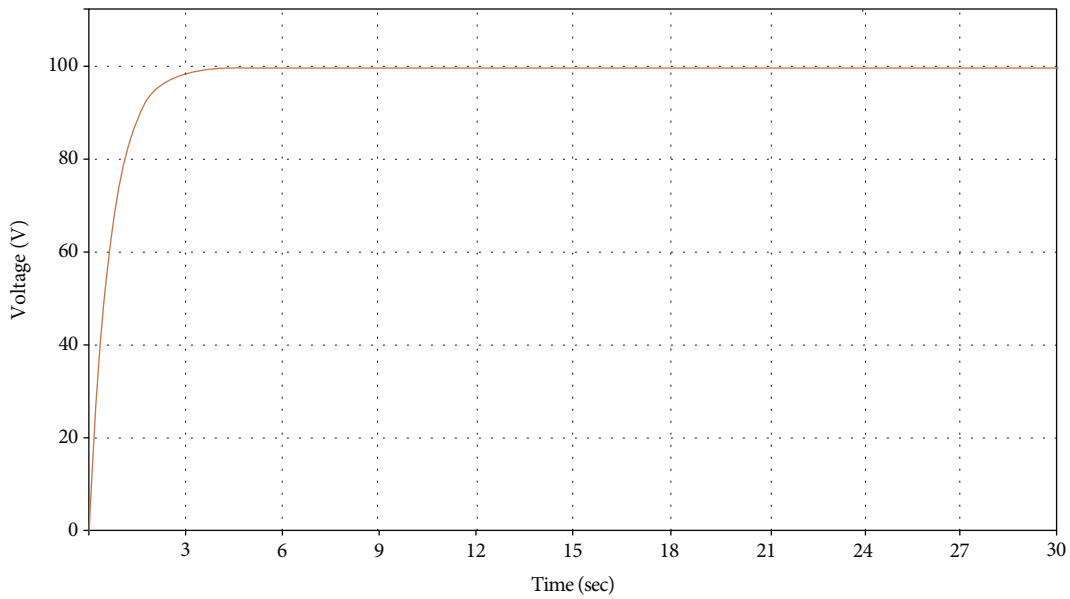
In Figures 10(a) and 10(b), the output values are great and the time to reach these values is short. However, the disadvantages of the boost converter are the fluctuated outputs. In Figure 10(a), the voltage oscillates from 350 to 450 VDC, and the current also oscillates as shown in Figure 10(b). Obviously, the output values of the boost converter have significant fluctuations that cause negative effects on power systems. In order to address this problem, the stabilizer system is used to generate a suitable voltage and current for the BTS system.

4.5. Stabilizer and Adapter System. The main functions of the stabilizer system are to create the desired voltage and current with a steady output to supply the base transceiver station (BTS) system. The stabilizer includes the buck converter and the controller. The buck converter is used to step down the voltage, but it is difficult to get to the desired output values without the controlling part. The controller is the determinant to exactly create the desired outputs by controlling MOSFET's switching of the buck converter.

By utilizing the stabilizer system, the 400 VDC voltage and 100 A current are stepped down to 48 VDC and 52 A,



(a) The output current



(b) The output voltage

FIGURE 9: The output of the solar cells.

as shown in Figure 11, and the time for reaching these output values is approximately one minute. The settling time of the stabilizer’s outputs is dependent on the output values of the boost converter. The higher the output values of the boost converter are produced, the shorter the settling time of the stabilizer’s outputs is taken.

4.6. Discussion. The amount of RF at the BTS is plentiful, and the generating energy is quite large to contribute to the system. The BTS power is supplied by the two main sources and two backup sources, which are the hybrid solar-RF energy, grid sources, generator, and batteries. Because the BTS system has some different sources, the supplying power system needs to be operated versatily to avoid interruption

in providing power to the loads. The supplying power system is divided into three modes depending on the source priority to supply power to loads. The hybrid energy source is the top priority and then the grid source. The batteries are only used when the above two sources are not available to meet power. During the normal mode, the first source choice is the hybrid energy system (HES), supplying power to the BTS system and charging batteries at the same time. When the HES is not sufficient to satisfy the power demand, the energy shortage will be supplied by the grid which is considered just a backup. During the unusual case, when the grid is not available and the HES is also not sufficient, the BTS power is supplied by the batteries. During battery usage, if the grid or the HES is available, the BTS system and batteries are

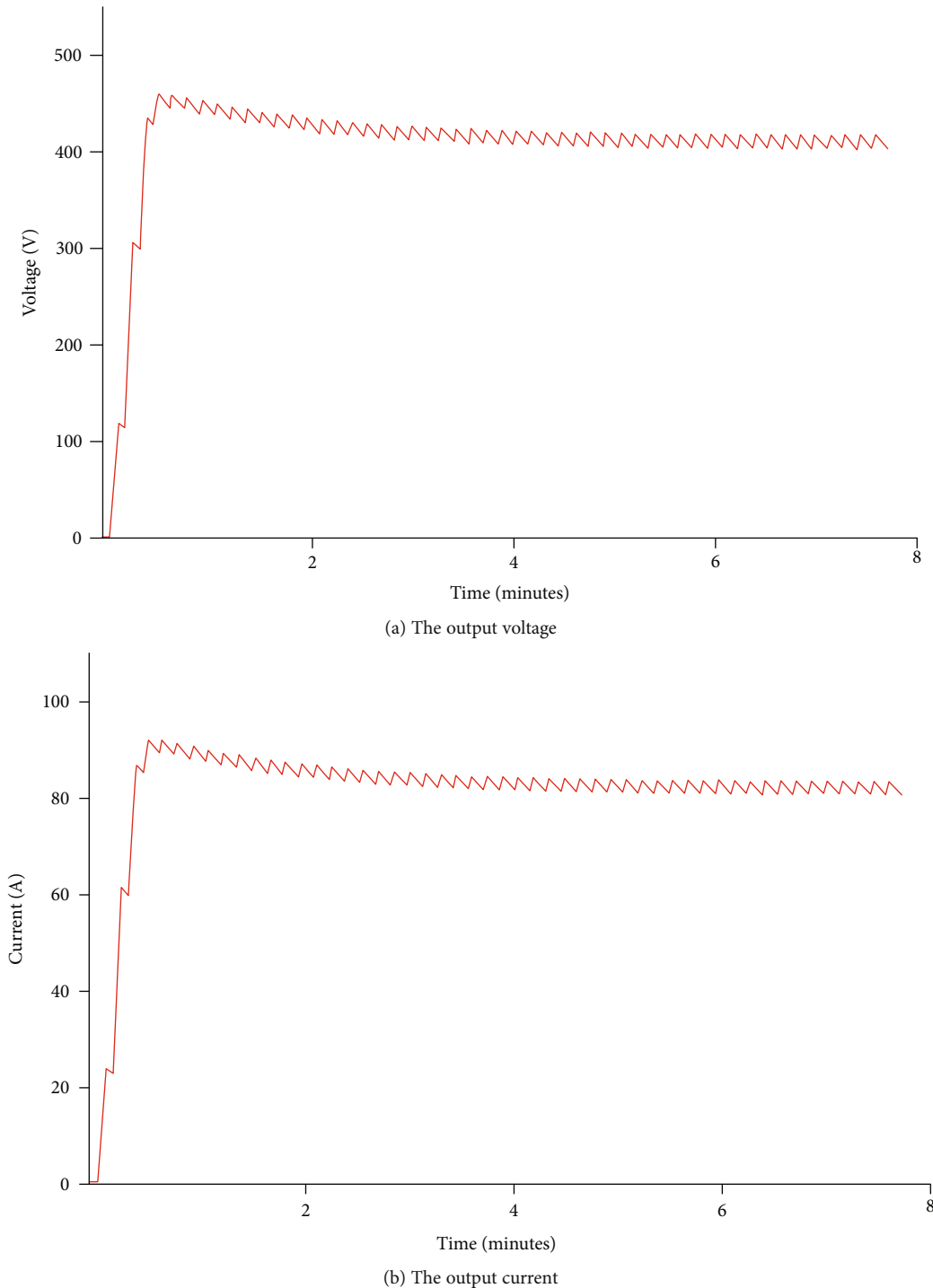


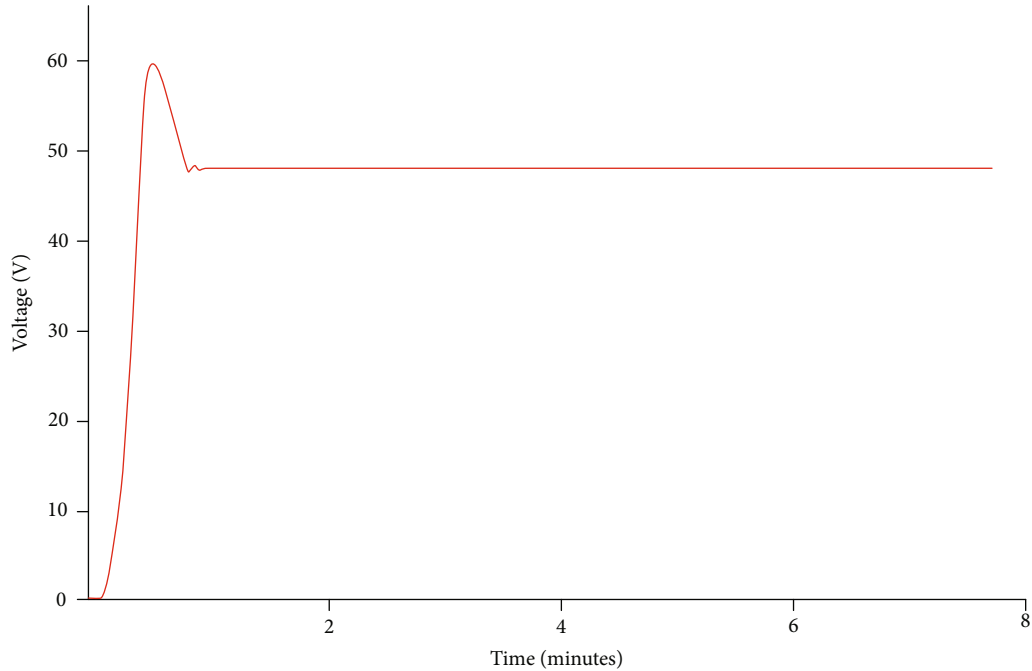
FIGURE 10: The output of the boost converter.

provided by the grid or the HES. In the last case, if the batteries run out, the generator is used to supply power for the system.

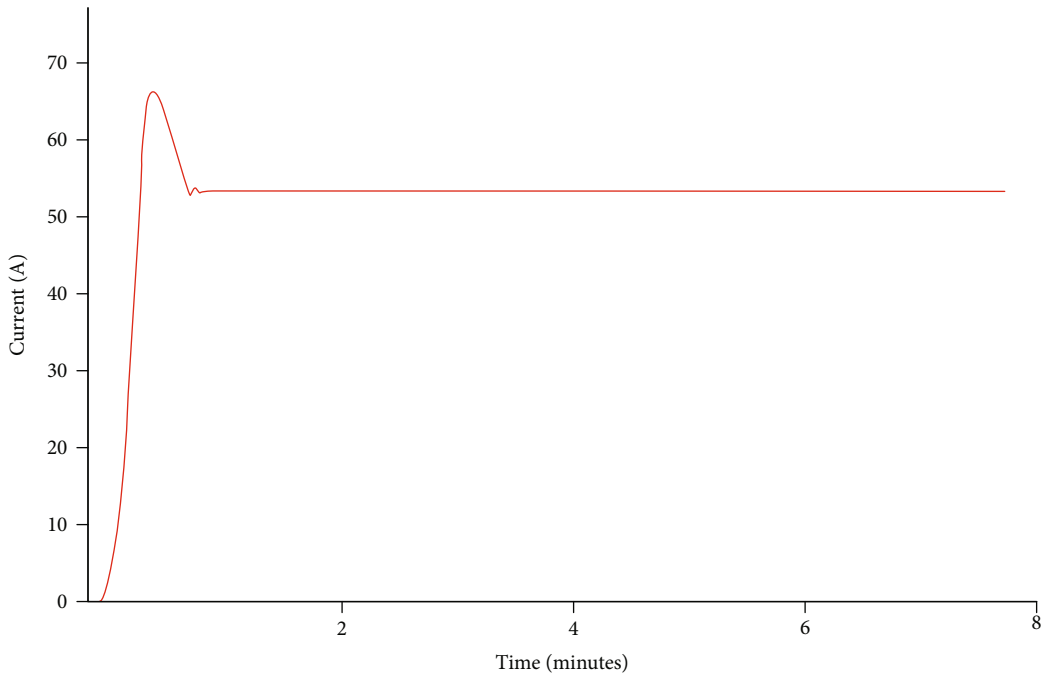
5. Conclusions and Future Work

In this work, we propose a new hybrid energy harvesting system for a specific purpose such as powering the base stations in communication networks. The hybrid solar-RF energy sys-

tem is designed, simulated, and calculated to evaluate the outcomes. It is shown that the proposed system can supply 52 A and 48 VDC to the BTS that would be sufficient. The hybrid system powering 2.5 kW power helps mitigate the huge power supplied by the grid. Therefore, the proposed power system contributes positively to both the environment and economic aspects. The results show promise that it could be practically deployed to relax the burden of energy supplies. This also



(a) The output voltage



(b) The output current

FIGURE 11: The output of the stabilizer.

supports mobile BTS working in different fields where need for communication capacity is increasing.

Since the management of the source power has not been mentioned yet, that would be the next steps to improve the systems. Besides, it is possible to improve the efficiency of the boost converter and buck converter by utilizing the fuzzy and particle swarm optimization methods to control the duty cycle of the MOSFET. Last, we are going to alleviate power dissipa-

tion during the on and off modes of the MOSFET by changing the hard switching to soft switching. A soft-switched device turns on and off at zero or nearly zero voltage or current.

Data Availability

Data used in this study are all in the manuscript. We are willing to provide all the data if needed.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Thai Nguyen University of Technology (TNUT), Vietnam, for the support.

References

- [1] N. Faruk, M. Y. Muhammad, O. Bello, A. Abdulkarim, A. John, and M. Gumel, "Energy conservation through site optimization for mobile cellular systems (base transceivers station optimization)," *Epistemics in Science, Engineering and Technology*, vol. 2, no. 1, pp. 26–33, 2012.
- [2] K. David, D. Dixit, and N. Jefferies, "2020 vision," *IEEE Vehicular Technology Magazine*, vol. 5, no. 3, pp. 22–29, 2010.
- [3] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, and J. Zhang, "Rethinking energy efficiency models of cellular networks with embodied energy," *IEEE Network*, vol. 25, no. 2, pp. 40–49, 2011.
- [4] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: a survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 4, pp. 524–540, 2011.
- [5] A. Spagnuolo, A. Petraglia, C. Vetromile, R. Formosi, and C. Lubritto, "Monitoring and optimization of energy consumption of base transceiver stations," *Energy*, vol. 81, pp. 286–293, 2015.
- [6] M. T. Nguyen, "Energy harvesting in wireless sensor networks: benefits and challenges," *ICSES Interdisciplinary Transactions on Cloud Computing, IoT, and Big Data*, vol. 4, no. 1, pp. 1–3, 2020.
- [7] F. K. Shaikh and S. Zeadally, "Energy harvesting in wireless sensor networks: a comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 1041–1054, 2016.
- [8] V. T. Vu, T. V. Quyen, L. H. Truong, A. M. Le, C. V. Nguyen, and M. T. Nguyen, "Energy efficient approaches in wireless sensor networks," *ICSES Transactions on Computer Networks and Communications*, vol. 6, no. 1, pp. 1–10, 2020.
- [9] C. Van Nguyen, T. Van Quyen, A. M. Le, L. H. Truong, and M. T. Nguyen, "Advanced Hybrid Energy Harvesting Systems for Unmanned Ariel Vehicles (UAVs)," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 1, pp. 34–39, 2020.
- [10] M. T. Nguyen, C. V. Nguyen, L. H. Truong et al., "Electromagnetic field based wpt technologies for uavs: a comprehensive survey," *Electronics*, vol. 9, no. 3, p. 461, 2020.
- [11] C. V. Nguyen, M. T. Nguyen, T. V. Quyen et al., "Hybrid solar-RF energy harvesting systems for electric operated wheel-chairs," *Electronics*, vol. 9, no. 5, p. 752, 2020.
- [12] S. A. Chowdhury and S. Aziz, "Solar-diesel hybrid energy model for base transceiver station (bts) of mobile phone operators," in *2nd International Conference on the Developments in Renewable Energy Technology (ICDRET 2012)*, pp. 1–6, Dhaka, Bangladesh, 2012.
- [13] T. V. Quyen, C. V. Nguyen, A. M. Le, and M. T. Nguyen, "Optimizing hybrid energy harvesting mechanisms for UAVs," *EAI Endorsed Transactions on Energy Web*, vol. 7, no. 28, 2020.
- [14] P. Sanjeevikumar and K. Rajambal, "Extra-high-voltage DC-DC boost converters topology with simple control strategy," *Modelling and simulation in Engineering*, vol. 2008, Article ID 593042, 8 pages, 2008.
- [15] H. Endo, T. Yamashita, and T. Sugiura, "A high-power-factor buck converter," in *PESC '92 Record. 23rd Annual IEEE Power Electronics Specialists Conference*, pp. 1071–1076, Toledo, Spain, 1992.
- [16] K. K. A. Devi, N. M. Din, and C. K. Chakrabarty, "Optimization of the voltage doubler stages in an rf-dc convertor module for energy harvesting," *Circuits and Systems*, vol. 3, no. 3, pp. 216–222, 2012.
- [17] T. R. Ansari, A. Khan, and I. Ansari, "Wireless charging of mobile battery via optimization of RF energy harvesting system," *International Journal of Scientific & Engineering Research*, vol. 6, no. 7, pp. 942–949, 2015.
- [18] M. Park and I.-K. Yu, "A study on the optimal voltage for mppt obtained by surface temperature of solar cell," in *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004*, vol. 3, pp. 2040–2045, Busan, South Korea, 2004.
- [19] G. Walker et al., "Evaluating mppt converter topologies using a matlab pv model," *Journal of Electrical & Electronics Engineering, Australia*, vol. 21, no. 1, p. 49, 2001.
- [20] J. L. Santos, F. Antunes, A. Chehab, and C. Cruz, "A maximum power point tracker for pv systems using a high performance boost converter," *Solar Energy*, vol. 80, no. 7, pp. 772–778, 2006.

Research Article

Association Game for Conflict Resolution between UEs and Small Cells

Jaesung Park 

School of Information Convergence, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Republic of Korea

Correspondence should be addressed to Jaesung Park; jaesungpark@kw.ac.kr

Received 19 February 2020; Revised 15 May 2020; Accepted 25 May 2020; Published 9 June 2020

Academic Editor: Hideyuki Takahashi

Copyright © 2020 Jaesung Park. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Small cells are expected to increase the capacity and the coverage of a wireless network. Since the transmission ranges of the small cells overlap with each other, a user equipment (UE) can be located within the service areas of multiple cells. Therefore, the association decisions made by the small cells and the UEs influence the energy efficiency of the cells and the data rates provided to the UEs. However, if they attempt to achieve their goals selfishly, the network may become unstable in that a UE has to change its associated cell and a cell has to change the set of UEs associated with it frequently. To tackle the problem, in this paper, we propose an association game between UEs and cells. We cast the problem as a matching game where each player has preference over the other player. Then, we propose an association algorithm composed of a UE-side part and a cell-side part. Through simulation studies, we show that the proposed method outperforms the other methods in that it can enhance the average energy efficiency of the small cell network and the data rates that UEs can obtain from the network.

1. Introduction

Data traffic exchanged through cellular networks has increased drastically over the recent years. According to Ericsson mobility report [1], mobile data traffic grows 82 percent between the first quarter in 2018 and the first quarter in 2019, and it is expected that the mobile data will grow 30 percent annually until 2024. In addition, the proportion taken by video traffic is also expected to increase. Video traffic took 60 percent of the 28 EB mobile data traffic per month in 2018, and it is forecasted that video traffic will take 74 percent of the 131 EB data traffic per month by 2024.

To accommodate the traffic growth and data hungry applications, cellular networks have been evolved by increasing the capacity and the coverage. For example, ITU-R specified the enhanced mobile broadband (eMBB) as one of the important 5G and beyond cellular network service categories. To increase the capacity of wireless network, various methods have been devised (for example, new radio [2], advanced interference management [3], and network slicing [4]). Among those, small cells are expected to increase both

the capacity and the coverage of the network in a cost and energy-effective manner [5, 6]. By deploying small cells densely, the downlink data rate provided to each UE (User Equipment) will be enhanced. However, as the number of small cells increases, the energy consumed by the network increases accordingly. Thus, energy efficient operation becomes one of the important design requirements of a small cell network [7].

Cell sleep methods have been proposed to increase the energy efficiency (EE) of a small cell network [8, 9]. The basic idea of the cell sleep is to switch off small cells when they are not necessary to provide reasonable services to the UEs having data to send. After determining cells to sleep, the cell sleep method forces the UEs connected to the sleeping cell to associate with new active cells. Power control has also been used widely to increase the EE of a network [10, 11]. The signal quality increases as the transmission power increases. However, the amount of interferences also increases with the transmission power, which results in decreasing the throughput of a network. Thus, the dynamic power control method seeks to find the optimal transmission power for

each cell in the network so that the energy efficiency of the network is maximized.

On the other hand, the purpose of UEs is to increase the service rates received from the network. In a small cell network, a UE is within the transmission ranges of multiple cells. Thus, the cell association method will significantly influence the data rates received by a UE. Conventionally, a UE monitors the signal strengths from neighboring cells and attempts to associate with the cell giving the highest signal strength. However, since a UE is not aware of the load condition of a cell, the cell giving the highest signal strength to a UE is not necessarily the one that can provide the highest data rate to the UE. The authors in [12] proposed an association method to balance the loads of the cells in the network. An interference model is proposed in [13] to predict the cell load. Using the estimated load, the authors proposed a joint optimization algorithm for user association and antenna tilt adjustment.

Through association management, a cell tries to enhance its EE while UEs attempt to maximize the data rate received from a network. Since the intention of the two entities (i.e., the cells and the UEs) may be conflicting, each entity may not be able to accomplish their goals if a cell and a UE try to achieve their goals selfishly. Therefore, the small cell network can be unstable in that UEs and cells keep changing their associations.

To resolve the issue, we propose a matching game between UEs and cells. UEs and cells have the most wanted counterpart to achieve their goals. Therefore, we cast the association problem as a matching problem between UEs and cells who have preferences over the others. Then, we devise an association algorithm that results in stable pairs of UEs and cells by matching each UE to its most preferred achievable cell. Through simulation studies, we show that the proposed method can increase both the EEs of the cells and the downlink data rate provided to UEs, compared with those acquired by the conventional methods.

The rest of the paper is organized as follows: after, we formulate the problem by describing the system model in Section 2; we present the proposed algorithm in Section 3. We evaluate the proposed method through simulation studies in Section 4 and conclude the paper in Section 5.

2. System Model

We consider a wireless network where small cells are deployed densely. Before we proceed, we present the notations used in this paper in Table 1.

We denote by C a set of small cells in the network. We assume that small cells are using the OFDMA and the time-frequency resources that a cell $i \in C$ can allocate to UEs are divided into the physical resource blocks (PRBs). We denote by B_p the bandwidth of a PRB and by B_M the maximum number of PRBs that a cell has. We assume that the frequency reuse factor in the network is one. We denote the set of UEs deployed in the network as U . We also denote by U_i the set of UEs associated with a cell i and by C_u the set of cells that a UE u can make an association with.

TABLE 1: Notations used.

Notations	Meaning
U_i	Set of UEs associated with a cell i .
C_u	Set of cells that a UE u can make an association with.
B_p	Bandwidth of a physical resource block.
B_M	The maximum number of a physical resource block in a cell.
$\beta_{u,i}$	SINR between a UE u and a cell i .
$\gamma_{u,i}$	Downlink data rate provided to a UE u by a cell i .
Γ_i	Total data rate that a cell i can provide.
E_i	Total energy consumed by a cell i .
ρ_i	Load of a cell i .
Ψ_i	Energy efficiency of a cell i .
Ω	Total energy efficiency of a system.

Let us consider the situation where a UE u is associated with a cell i . If the transmission power of a cell i is p_i , the power received at u becomes

$$\alpha_{u,i} = p_i h_{u,i}, \quad (1)$$

where $h_{u,i}$ is the path gain between u and i . The path gain includes all the factors that can influence the received signal power such as path loss, shadowing, and antenna height. Then, the SINR (signal-to-interference-plus-noise-ratio) between u and i is expressed as

$$\beta_{u,i} = \frac{\alpha_{u,i}}{\sum_{j \in C - \{i\}} \alpha_{u,j} + N_0}, \quad (2)$$

where N_0 is a noise power.

The data rate that a cell i can provide to a UE u depends on the amount of resources that i allocates to u and the SINR between them. We denote by $m_{u,i}$ the number of PRBs that i allocates to u . Then, by the Shannon's capacity formula, the downlink data rate provided to u by i is given as

$$\gamma_{u,i} = m_{u,i} B_p \log_2(1 + \beta_{u,i}). \quad (3)$$

$m_{u,i}$ depends on the scheduler used by a cell i and the number of UEs being served by i . Since we focus on the association management method in this paper, we assume that each cell serves the UEs associated with it in a round-robin manner. Then,

$$m_{u,i} = \frac{B_M}{|U_i|}, \quad (4)$$

where $|U_i|$ is the cardinality of the set U_i . However, we note that the type of the scheduler does not influence the operation of the proposed method.

From Equation (3), the total data rate that a cell can provide is obtained as

$$\Gamma_i = \sum_{u=1}^{|U_i|} \gamma_{u,i} = \sum_{u=1}^{|U|} I_{u,i} \gamma_{u,i}, \quad (5)$$

where $I_{u,i}$ is the indicator function such that

$$I_{u,i} = \begin{cases} 1, & u \in U_i, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

The energy consumption model of a cell is presented in [14]. The model is composed of two parts. The first part represents the amount of energy when a cell is not involved in transmitting or receiving packets while the second part represents the amount of power consumed by a cell which depends on the load of the cell. Specifically, the total energy consumed by a cell i is given as

$$E_i = n_A P_0 + \Delta \rho_i p_i, \quad (7)$$

where P_0 is the fixed amount of energy consumed by a cell, n_A is the number of antennas of a cell, Δ is the slope of the energy consumption, and ρ_i is the load of a cell.

The load of a cell is defined as the ratio of the amount of allocated PRBs to B_M . However, since a cell in the system allocates all the resources to UEs associated with it (Equation (4)), the load of a cell becomes 1 regardless of the number of UEs being served by the cell. To cope with the issue, we assume that there is the minimum downlink data rate γ_{th} that makes a UE keep its current association. To define ρ_i , we let $\tilde{m}_{u,i}$ be the minimum number of PRBs needed to satisfy $\gamma_{u,i} \geq \gamma_{th}$. From Equation (3), $\tilde{m}_{u,i}$ is given as

$$\tilde{m}_{u,i} = \left\lceil \frac{\gamma_{th}}{B_p \log_2(1 + \beta_{u,i})} \right\rceil, \quad (8)$$

where $\lceil x \rceil$ is the smallest integer that is not smaller than x . Then, the load of a cell is given as

$$\rho_i = \sum_{u=1}^{|U|} \frac{I_{u,i} \tilde{m}_{u,i}}{B_M}. \quad (9)$$

The energy efficiency of a cell is defined as the total data rate the cell provides over the amount of energy consumed to support the total data rate. From Equations (5) and (7), the energy efficiency of a cell is obtained as

$$\Psi_i = \frac{\Gamma_i}{E_i}. \quad (10)$$

Using Equation (10), we get the total energy efficiency of a system.

$$\Omega = \sum_{i=1}^{|C|} \sum_{u=1}^{|U|} \frac{I_{u,i} \gamma_{u,i}}{n_A P_0 + \Delta \rho_i p_i}. \quad (11)$$

We note that the set of associations between cells and UEs (i.e., $I_{u,i}$) influences Ψ_i and $\gamma_{u,i}$. In the respect of a cell, its purpose is to increase its energy efficiency. Thus, when a cell i receives the association requests from UEs, it attempts to accept the UEs who can increase its energy efficiency. On the contrary, the goal of a UE is to get the downlink data rate as high as possible. From Equation (3), we see that $\gamma_{u,i}$ depends on $\beta_{u,i}$ and $m_{u,i}$. As we can see in Equation (2), the association made between a UE and a cell affects $\beta_{u,i}$. The number of PRBs allocated to a UE by a cell is also influenced by the association made between the cell and UEs (see Equation (4)). Therefore, the set of optimal associations for a cell may not be the optimal association for a UE because their purposes are different. If each entity (a cell and a UE) attempts to achieve their goals selfishly, the associations made between cells and UEs in a network may not be stabilized because UEs and cells may keep changing their associations to accomplish their goals.

To resolve the conflict, we propose an association method using a matching game theory [15]. Specifically, by noting that each party has a preferred counterpart to accomplish its purpose, we consider the association problem as the matching with preference game with two sets of players C and U . Then, we devise a distributed algorithm that determines the association pairs between a set of small cells and a set of UEs.

3. Association Game

3.1. Matching Algorithm. In this section, we present our matching-based association management algorithm. The algorithm is composed of two parts. One part is performed in a UE, and the other part is operated in a cell. When a UE u needs to make an association with a cell, it measures $\beta_{u,i}$ from all the cells in C_u by monitoring the reference signals. After collecting $\beta_{u,i}$, u sorts the cells in C_u in a descending order according to $\beta_{u,i}$. We denote the sorted set of cells as \tilde{C}_u . Then, u attempts to associate with the first cell (say, i) in \tilde{C}_u by sending an Association Req. message to it. If the association request is accepted by i , u makes an association with i . If the request is rejected, u removes i from \tilde{C}_u and repeats the same process again. When u does not make an association with a cell until \tilde{C}_u becomes empty, it restarts the process by measuring $\beta_{u,i}$ again. When a UE u is currently associated with a cell i , the cell can force u to handover by sending a Cell Redirect message to u . In that case, u starts the association procedure stated above with $\tilde{C}_u - \{i\}$. We summarize the UE side operation in Algorithm 1.

The cell side algorithm operates as follows: when a cell i receives an Association Req. message from u at time t , i checks whether the data rate provided to each UE already associated with it before t (denoted by $U_i(t-1)$) can still be larger than γ_{th} if it accepts the association request from u . Specifically, given $\tilde{U}_i(t) = U_i(t-1) \cup \{i\}$, i calculates the

```

UE side algorithm.
1  Cell Search:
2   measure  $\beta_{u,i} \forall i \in C_u$ 
3   sort  $i \in C_u$  according to  $\beta_{u,i}$  in a descending order to make  $\tilde{C}_u$ 
4  Association Attempt:
5   while  $\tilde{C}_u \neq \emptyset$  do
6     send Association Req. to the first cell in  $\tilde{C}_u$ 
7     if receive Accept Association Req. then
8       make an association with the cell
9     else
10       $\tilde{C}_u = \tilde{C}_u - \{i\}$ 
11    if  $\tilde{C}_u == \emptyset$  and Not Associated then
12      go to Cell Search
13  Cell Reselection:
14  if receive Cell Redirect from a cell  $i$  then
15    go to Association Attempt with  $\tilde{C}_u = \tilde{C}_u - \{i\}$ 

```

ALGORITHM 1:

number of PRBs allocated to a UE v in $\tilde{U}_i(t)$ using Equation (4) as $\tilde{m}_{v,i}(t) = B_M / |\tilde{U}_i(t)|$. Then, the cell i obtains the downlink data rate provided to each UE v in $\tilde{U}_i(t)$ as $\hat{\gamma}_{u,i}(t) = \tilde{m}_{v,i}(t) B_P \log_2(1 + \beta_{v,i})$. After obtaining $\hat{\gamma}_{u,i}(t)$, i compares $\hat{\gamma}_{u,i}(t)$ with γ_{th} . If $\hat{\gamma}_{u,i}(t) > \gamma_{th}$, for all $v \in \tilde{U}_i(t)$, it means that the cell does not drive any UE in $\tilde{U}_i(t)$ to change its serving cell even though i accepts u . Then, i checks whether its energy efficiency can increase by accommodating u .

If a cell i accepts the association request, the load of the cell becomes $\tilde{\rho}_i(t) = \sum_{v \in \tilde{U}_i(t)} \tilde{m}_{v,i} / B_M$. If we denote by $\tilde{\Psi}_i(t)$ the energy efficiency of a cell i when it accepts the association request from u , using Equation (10), i gets

$$\tilde{\Psi}_i(t) = \frac{\sum_{v \in \tilde{U}_i(t)} \hat{\gamma}_{v,i}}{n_A P_0 + \Delta \tilde{\rho}_i(t) P_i}. \quad (12)$$

If $\tilde{\Psi}_i(t) > \Psi_i(t-1)$, i accepts the association request from u to increase its energy efficiency. On the other hand, if $\tilde{\Psi}_i(t) \leq \Psi_i(t-1)$, the energy efficiency of i reduces if i accepts u . However, instead of rejecting the association request from u immediately, i checks whether its energy efficiency can be increased by releasing an association with a UE in $U_i(t-1)$ and accepting u . Specifically, let us denote $U_{i,v}(t) = U_i(t-1) \cup \{u\} - \{v\}$ ($v \in U_i(t-1)$). Then, for all $v \in U_i(t-1)$, i obtains its energy efficiency when the set of UEs associated with it is $U_{i,v}(t)$ as follows.

$$\Psi_{i,v}(t) = \frac{\sum_{k \in U_{i,v}(t)} \gamma_{k,i}}{n_A P_0 + \Delta \rho_{i,v} P_i}, \quad (13)$$

where $\rho_{i,v}$ is the load of a cell when the set of UEs associated with i is $U_{i,v}(t)$. A cell i searches for a UE v^* that can maximize the $\Psi_{i,v}(t)$. In other words,

$$v^* = \arg \max_{v \in U_i(t-1)} \Psi_{i,v}(t). \quad (14)$$

Thus, the maximum energy efficiency that the cell i can achieve by replacing a UE in $U_i(t-1)$ with u becomes $\Psi_{i,v^*}(t)$. Then, the condition $\Psi_{i,v^*}(t) > \Psi_i(t-1)$ means that the energy efficiency of a cell i at time $t-1$ (i.e., $\Psi_i(t-1)$) is smaller than the energy efficiency that i can get by releasing an association with v^* and accepting the association request from u . Thus, a cell i accepts the association request from u and releases an association with v^* by sending a Cell Redirect message to v^* . On the contrary, if $\Psi_{i,v^*}(t) \leq \Psi_i(t-1)$, i cannot get a gain in terms of its energy efficiency. Therefore, i rejects the association request by sending a Reject Association Req. to u . We show the algorithm operating in a cell in Algorithm 2.

3.2. Stability Analysis. To prove the stability of the proposed algorithm, we introduce the following definition.

Definition 1. A member x in one set said to be acceptable to a member y in the other set if y prefers x at least as well as to be not associated at all.

We state the stability of the proposed algorithm with the following proposition.

Proposition 2. The set of associations between the UEs in U and the cells in C is a stable matching μ_s in that all the UEs and cells are associated with their mutually acceptable counterparts.

Proof. Let us consider the case that the matching μ_s achieved by the proposed method does not associate a UE u with a cell i , and i prefers u to $v \in U_i$. It means either i is not acceptable to u , u did not attempt to make an association with i . If i rejects the association request from u , it negates the statement that i prefers u to $v \in U_i$. Thus, μ_s does not associate i with a UE to whom i is not acceptable. In addition, μ_s associates i with v acceptable to i .

```

Cell side algorithm.
1  Receive Association Req. from  $u$ :
2   $\tilde{U}_i(t) = U_i(t-1) \cup \{u\}$ 
3   $\hat{m}_{v,i}(t) = B_M / |\tilde{U}_i(t)|$ 
4   $\hat{\gamma}_{u,i}(t) = \hat{m}_{v,i}(t) B_P \log_2(1 + \beta_{v,i})$ 
5  if  $\hat{\gamma}_{u,i}(t) > \gamma_{th}$ ,  $\forall v \in \tilde{U}_i(t)$  then
6   $\tilde{\rho}_i(t) = \sum_{v \in \tilde{U}_i(t)} \hat{m}_{v,i} / B_M$ 
7   $\tilde{\Psi}_i(t) = \sum_{v \in \tilde{U}_i(t)} \hat{\gamma}_{v,i} / n_A P_0 + \Delta \tilde{\rho}_i(t) p_i$ 
8  if  $\tilde{\Psi}_i(t) > \Psi_i(t-1)$ , then
9    Send Accept Association Req. to  $u$ 
10 else
11   $U_{i,v}(t) = U_i(t-1) \cup \{u\} - \{v\}$ 
12  get  $\Psi_{i,v}(t) = \sum_{k \in U_{i,v}(t)} \gamma_{k,i} / n_A P_0 + \Delta \rho_{i,v} p_i, \forall v \in U_i(t-1)$ 
13  find  $v^* = \text{argmax}_{v \in U_i(t-1)} \Psi_{i,v}(t)$ 
14  if  $\Psi_{i,v^*}(t) > \Psi_i(t-1)$  then
15    send Accept Association Req. to  $u$ 
16    send Cell Redirect to  $v^*$ 
17 else
18    send Reject Association Req. to  $u$ 

```

ALGORITHM 2:

TABLE 2: Simulation parameters.

Parameters	Values
Simulation area (Y)	$100 \times 100 \text{ m}^2$
Density of small cells	$200/Y$
System bandwidth	5 MHz
Bandwidth of a PRB	180 KHz
Max. number of PRBs in a cell	25
Tx power of a cell	21 dBm
Antenna gain of a cell	3.0 dBi
Antenna gain of a UE	2.0 dBi
UE scheduling policy	Round-Robin
n_A	1
P_0	6.8 W
Slope of energy consumption	4
Path loss model	$140.7 + 36.7 \log(d)$
Noise power	-111.45 dBm
Shadowing model	Lognormal $\sim N(0, 8 \text{ dB})$
γ_{th}	256 Kbps

Let us consider the other case that u and i are not associated with by the stable matching μ_s , and u prefers i to the cell j that it is currently associated with. Since u likes i more than j , i must be acceptable to u . Moreover, u must have attempted to be associated with i before it tried j . Since μ_s does not associate u with i , it means that i rejected the association request from u and accepted the association request from other UE v who is acceptable and at least as preferable as u to i . Thus, μ_s associates a cell i with a UE acceptable to i . Since μ_s associates u with j , u and j are mutually acceptable to each other. There-

TABLE 3: Proportion of cells whose energy efficiencies are less than x .

λ_{ue}	x	0.5 Mbps/W	1 Mbps/W	4 Mbps/W
200/ Y	Proposed	0.53	0.69	0.96
	SSBM	0.53	0.72	0.98
	DRBM	0.53	0.72	0.98
400/ Y	Proposed	0.32	0.52	0.93
	SSBM	0.33	0.57	0.97
	DRBM	0.35	0.57	0.97
800/ Y	Proposed	0.13	0.35	0.92
	SSBM	0.16	0.49	0.98
	DRBM	0.20	0.50	0.99

fore, in this case, the UEs and the cells are associated with their mutually acceptable counterparts.

If a cell i is not acceptable to a UE u , u would stay unassociated rather than be associated with i . Thus, u will not be associated with i by the μ_s . Likewise, if u is not acceptable to i , i will reject the association request if it receives one from u . Therefore, μ_s does not make an association between a UE and a cell if they are not mutually acceptable.

Since UEs and cells are associated with their mutually acceptable counterparts, they do not change the current association, which makes the small cell network stable.

4. Simulation Results and Discussions

In this section, we compare the performance of the proposed method with that of two alternative association methods under the same simulation environments. To make the

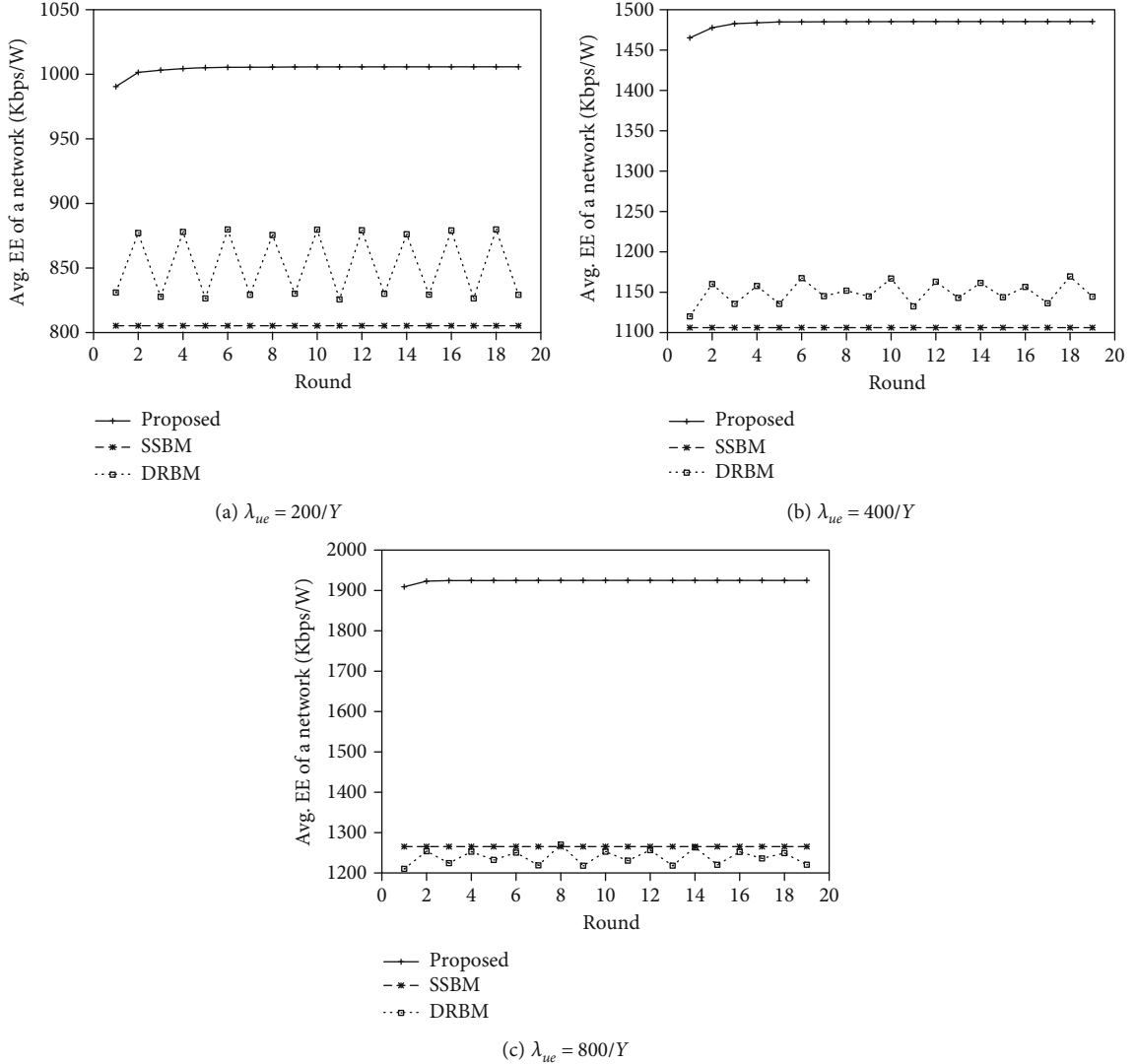


FIGURE 1: Comparison of the average network energy efficiency.

performance comparison, we make an event-driven simulator by using C language. The simulator is implemented in a cygwin on a windows 10 Pro for workstations. To run the simulator, we use a Dell precision 7920 workstation which has Intel Xeon(R) gold 5118 12C/24T CPU at 2.3 GHz and 16GB 2,666 MHz DDR4 RAM.

The first alternative method uses $\beta_{u,i}$. A UE u asks for an association to a cell i who gives u the highest SINR, and i accepts the request unconditionally. The method is widely used in many wireless networks. Henceforth, we name the first alternative as SSBM (Signal-Strength-Based Method). The second association method attempts to maximize the downlink data rate provided to each UE. In this method, a UE makes an association request to a cell that can give it the highest $\gamma_{u,i}$ and a cell accepts any association request unconditionally. In this method, a UE keeps changing its associated cell whenever it finds another cell that can give it a higher data rate than that it currently receives. Hereafter, we will call the second method as DRBM (Downlink-Rate-Based Method).

We deployed small cells in $100 \times 100 \text{ m}^2$ region according to the homogeneous point Poisson process with mean λ_{sc} . We denote by Y the simulation area. We also deployed UEs in the same region according to the homogeneous point Poisson process with mean λ_{ue} . The small cell network is configured according to [16, 17]. We set the system bandwidth to 5 MHz and $B_p = 180$ KHz. Thus, each small cell has $B_M = 25$ PRBs. The transmission power of a cell is set to 21 dBm. The antenna gain of a cell is configured as 3.0 dBi and that of a UE is set to 2.0 dBi. Each cell serves the UEs associated with it in a round-robin manner. We set the number of antennas in a cell (n_A) to 1, $P_0 = 6.8$ W, and $\Delta = 4$. The path loss is configured as $140.7 + 36.7 \log(d)$, where d is the distance between a cell and a UE. We also configure the noise power of -111.45 dBm and the log-normal shadowing with zero mean and variance of 8 dB. When our method is used, we set $\gamma_{th} = 256$ Kbps. After setting $\lambda_{sc} = 200/Y$, we observe how each method matches the set of UEs to the set of small cells by varying λ_{ue} . We summarize the simulation parameters in Table 2.

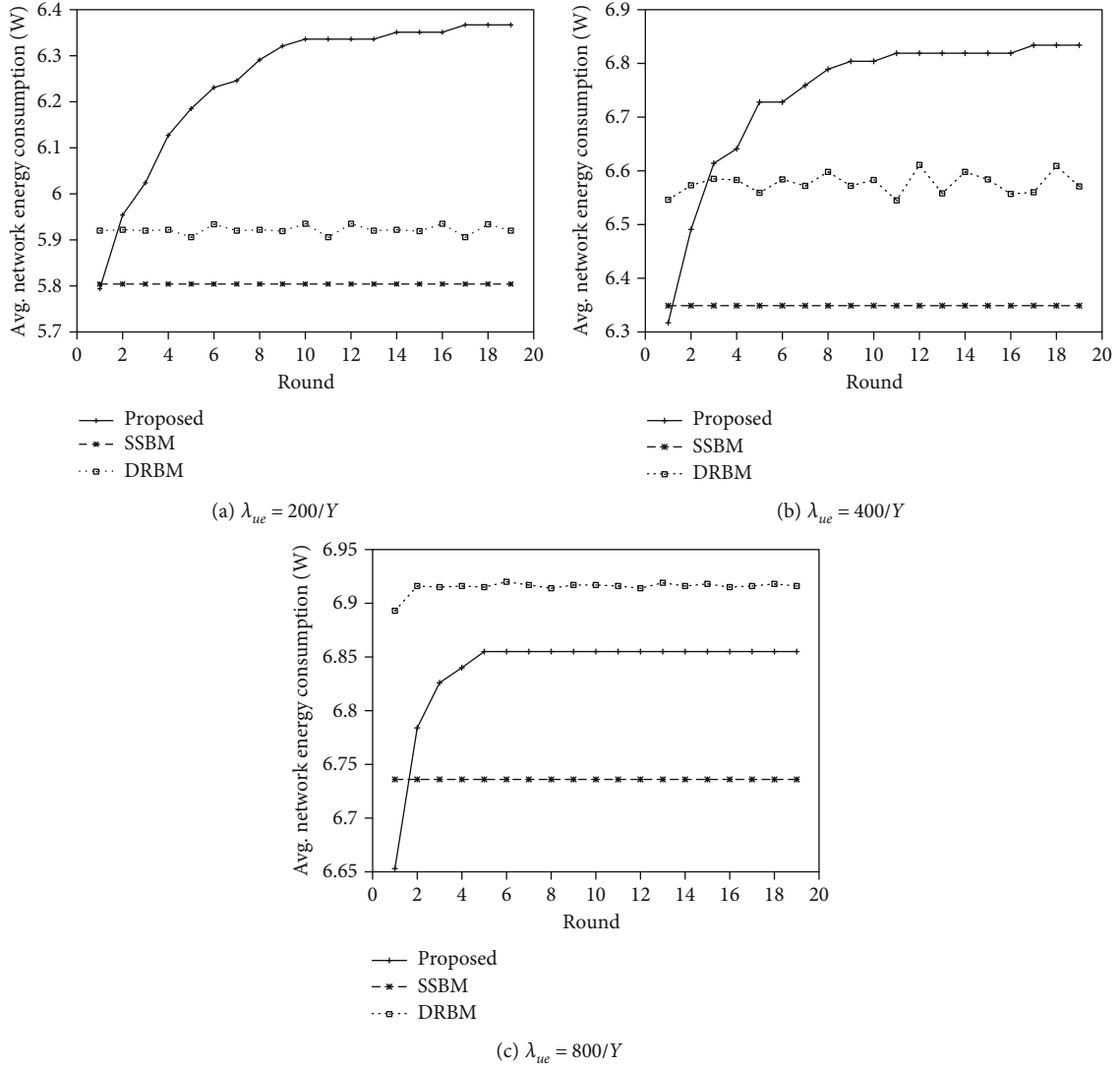


FIGURE 2: Comparison of the average network energy consumption.

We measure the energy efficiency of each cell ($\Psi_i, \forall i \in C$) and present the proportion of the cells whose energy efficiencies are less than x in Table 3. We can observe in the table that the energy efficiencies of the cells obtained by the proposed method are higher than those acquired when SSBM and DRBM are used. When $\lambda_{ue} = 400/Y$, 52% of the small cells have less than 1 Mbps/W of EE when the proposed method is used. On the contrary, the proportion is increased to 57% when SSBM and DRBM are used. In the table, we can also observe that the energy efficiency gap between the proposed method and the other methods increases as the number of UEs increases. For example, we inspect the proportion of the cells whose energy efficiency is less than 1 Mbps/W (denoted by ζ). When $\lambda_{ue} = 200/Y$, $\zeta = 0.72$ when SSBM and DRBM are used while the proposed method reduces ζ to 0.69. When λ_{ue} increases to $800/Y$, SSBM results in $\zeta = 0.49$ and DRBM gives $\zeta = 0.50$. On the contrary, the proposed method decreases $\zeta = 0.35$, which means that 75%

of the cells have Ψ_i s higher than 1 Mbps/W (approximately 25% increase compared with the other methods).

In Figure 1, we compare the average network energy efficiency, $\bar{\Omega} = \Omega/|C|$. In this figure, the x -axis represents the number of round. We define the round as the time interval for all the UEs in U which are not associated with any cell to make an association attempt once. In this figure, we can observe that the proposed method stabilizes the average network energy efficiency because it finds such pairs of UEs and cells that no UE in U and no cell in C want to change its associated counterpart. We can also observe that the proposed method results in higher network energy efficiency for all λ_{ue} , compared with SSBM and DRBM.

To scrutinize the results, we inspect the average energy consumption in the network ($\bar{E} = \sum_{i \in C} E_i/|C|$) and the average downlink data rate provided by the network ($\bar{T} = \sum_{i \in C} \Gamma_i/|C|$) in Figures 2 and 3, respectively. In Figure 2, we observe that \bar{E} obtained after the proposed method is stabilized is larger than

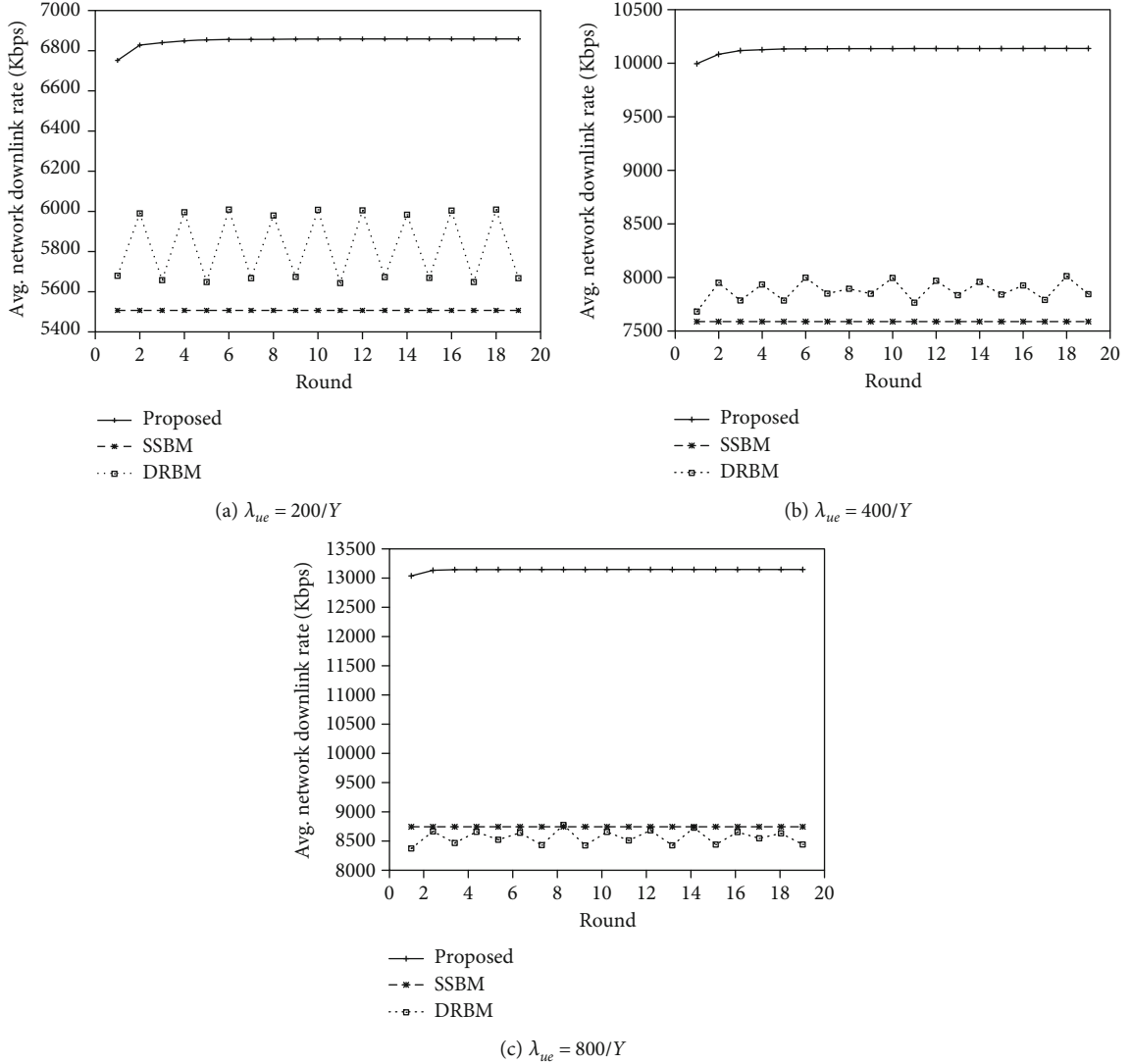


FIGURE 3: Comparison of the average downlink data rate provided by the network.

those of SSBM and DRBM when $\lambda_{ue} = 200/Y$ and $\lambda_{ue} = 400/Y$. However, when λ_{ue} increases to $800/Y$, \bar{E} of DRBM is larger than that of the proposed method. This is attributed to the fact that the proposed method distributes UEs to cells more evenly than the other methods. When SSBM and DRBM are used, UEs select their serving cells selfishly to increase their profits ($\beta_{u,i}$ or $\gamma_{u,i}$) and cells accept the association requests from UEs unconditionally. Therefore, it is highly likely that some cells are crowded with UEs while the other cells are lightly loaded [18]. On the contrary, in the case of the proposed method, cells attempt to increase their profits (i.e., energy efficiencies) by determining whether to accept the association requests from UEs or not. Since UEs cannot change their serving cells unilaterally, the problem that UEs select the same cell at the similar time to overload the cell is reduced by the proposed method.

Therefore, when $\lambda_{ue}/\lambda_{sc}$ is small (Figures 2(a) and 2(b)), the number of cells that are not associated with any UE (i.e., $U_i = \emptyset$) is the smallest when the proposed method is used. Since a cell that does not serve a UE consumes the

smallest energy and the UE density is small, \bar{E} obtained by the proposed method are the highest in these environments. As the UE density increases, $|U_i|$ increases on average. Therefore, as we can observe in Figure 2(c), \bar{E} produced by all the methods increase with λ_{ue} . However, the proposed method shows the smallest increase rate in \bar{E} , which makes \bar{E} obtained by the proposed method smaller than that of DRBM when $\lambda_{ue} = 800/Y$.

When we inspect \bar{T} in Figure 3, the proposed method outperforms the other methods. This is attributed to the fact that unlike the other methods, a cell using the proposed method resolves the contention among the UEs. The downlink data rate given to a UE depends not only on the SINR between a UE and a cell but also on the load of a cell serving the UE. Since SSBM does not consider cell loads, it is likely that some cells are highly congested while the other cells are lightly loaded, which results in the lowest average network downlink rate. When DRBM is used, a UE keeps changing its associated cell selfishly whenever it finds that the other

TABLE 4: Proportion of UEs whose $\gamma_{u,i}$ are less than x .

λ_{ue}	x	2 Mbps	5 Mbps	10 Mbps
200/Y	Proposed	0.32	0.58	0.81
	SSBM	0.29	0.68	0.87
	DRBM	0.27	0.67	0.87
400/Y	Proposed	0.32	0.59	0.81
	SSBM	0.46	0.80	0.92
	DRBM	0.44	0.80	0.92
800/Y	Proposed	0.29	0.61	0.83
	SSBM	0.67	0.90	0.97
	DRBM	0.65	0.91	0.98

cell can give it higher $\gamma_{u,i}$ than that it currently receives. Therefore, we can observe in Figure 3 that \bar{T} fluctuates over time. Unlike these methods where a cell accepts an association request unconditionally, a cell i using the proposed method checks γ_{th} provided to the UEs in U_i and Ψ_i when it decides whether to accept a new association request or not. Therefore, the proposed method can achieve higher \bar{T} than those of SSBM and DRBM by resolving the contention among UEs. Since the gain obtained in \bar{T} outweighs the loss occurred in \bar{E} , $\bar{\Omega}$ acquired by the proposed method becomes larger than those of the other methods.

We also inspect the proportion of UEs whose $\gamma_{u,i}$ are less than x . We denote such a proportion as $v(x)$. We measure $v(x)$ obtained by each method after the 15-th round and present them in Table 4 with different x . We observe that $v(x)$ obtained by the proposed method is lower than those acquired by the other methods for all x and λ_{ue} except $\lambda_{ue} = 200/Y$ and $x = 2$ Mbps. In this case, $v(2 \text{ Mbps})$ obtained by the proposed method is 0.32 which is higher than those obtained by the other methods. However, as λ_{ue} and x increase, the proposed method achieves smaller $v(2 \text{ Mbps})$ than the other methods. For example, when $\lambda_{ue} = 800/Y$ and $x = 5$ Mbps, only 61% of UEs receive less than 5 Mbps from the network when the proposed method is used. However, when SSBM and DRBM are used, 90% of UEs receive less than 5 Mbps. This is attributed to the fact that, as λ_{ue} increases, more UEs compete for the network resources. Since UEs act selfishly to increase their $\gamma_{u,i}$ when SSBM and DRBM are used, they cannot resolve the contention. Therefore, they are associated with a cell that cannot give them higher $\gamma_{u,i}$ that they can get through a matching game. We also observe in the table that the difference between the $v(x)$ obtained by the proposed method and those acquired by SSBM and DRBM becomes larger as the number of UEs in the network increases.

5. Conclusions and Future Works

In this paper, we proposed an association algorithm that matches a set of UEs to a set of cells. Since the proposed algorithm considers $\gamma_{u,i}$, it can increase the proportion of UEs who receive higher downlink data rates from the network,

compared with the SSBM and DRBM. In addition, the proposed method enhances the proportion of cells who have higher energy efficiency than those obtained by SSBM and DRBM because it also considers Ψ_i when the association is made. As a future work, we are planning to expand the association method to satisfy the requirements imposed on the small cell networks to serve massive IoT devices.

Data Availability

All the data used in this manuscript are included in the manuscript.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Research Resettlement Fund for the new faculty of Kwangwoon University in 2019 and by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (NRF-2018R1D1A1B07050893).

References

- [1] Ericsson, *Ericsson Mobility Report*, Ericsson, 2019.
- [2] M. G. Kibria, K. Nguyen, G. P. Villardi, K. Ishizu, and F. Kojima, "Next generation new radio small cell enhancement: architectural options, functionality and performance aspects," *IEEE Wireless Communications*, vol. 25, no. 4, pp. 120–128, 2018.
- [3] C. D. Nwankwo, L. Zhang, A. Qudus, M. A. Imran, and R. Tafazolli, "A survey of self-interference management techniques for single frequency full duplex systems," *IEEE Access*, vol. 6, pp. 30242–30268, 2018.
- [4] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: a survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.
- [5] F. Al-Turjman, E. Ever, and H. Zahmatkesh, "Small cells in the forthcoming 5G/IoT: traffic modelling and deployment overview," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 28–65, 2019.
- [6] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: current issues and machine learning-assisted solutions," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 1, pp. 426–471, 2020.
- [7] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.
- [8] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, 2016.
- [9] R. Tao, W. Liu, X. Chu, and J. Zhang, "An energy saving small cell sleeping mechanism with cell range expansion in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 5, pp. 2451–2463, 2019.

- [10] J. Zheng, Y. Wu, N. Zhang, H. Zhou, Y. Cai, and X. Shen, "Optimal power control in ultra-dense small cell networks: a game-theoretic approach," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4139–4150, 2017.
- [11] S. Mishra and C. S. R. Murthy, "Increasing energy efficiency via transmit power spreading in dense Femto cell networks," *IEEE Systems Journal*, vol. 12, no. 1, pp. 971–980, 2018.
- [12] Y. Wang, S. Chen, H. Ji, and H. Zhang, "Load-aware dynamic biasing cell association in small cell networks," *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 2684–2689, Sydney, NSW, Australia, 2014.
- [13] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1974–1988, 2013.
- [14] G. Auer, V. Giannini, C. Desset et al., "How much energy is needed to run a wireless network?," *IEEE Wireless Communications Magazine*, vol. 18, no. 5, pp. 40–49, 2011.
- [15] A. E. Roth and M. Sotomayor, "Chapter 16 Two-sided matching," *Handbook of Game Theory with Economic Applications*, vol. 1, 1992.
- [16] 3GPP, *Radio Frequency (RF) System Scenarios (Release 15)*, 3GPP TR 36.942 V15.0.0, The 3rd Generation Partnership Project (3GPP), 2018.
- [17] 3GPP, *Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, 3GPP TR 36.814 V9.2.0, The 3rd Generation Partnership Project (3GPP), 2018.
- [18] Q. Ye, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

Research Article

Analysis of Outage Probability and Throughput for Energy Harvesting Full-Duplex Decode-and-Forward Vehicle-to-Vehicle Relay System

Ba Cao Nguyen ¹, Nguyen Nhu Thang,¹ Tran Manh Hoang,¹ and Le The Dung ^{2,3}

¹Telecommunications University, Khanh Hoa Province, Vietnam

²Division of Computational Physics, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

³Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Correspondence should be addressed to Le The Dung; lethedung@tdtu.edu.vn

Received 5 January 2020; Revised 15 April 2020; Accepted 8 May 2020; Published 29 May 2020

Academic Editor: Gianluigi Ferrari

Copyright © 2020 Ba Cao Nguyen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we evaluate the performance of a vehicle-to-vehicle (V2V) system where full-duplex relay (FDR) harvests the energy from source and uses decode-and-forward (DF) protocol to forward data from source to destination. Unlike existing works about FDR systems, we consider the scenario that both relay and destination are moving vehicles, leading to the channel between relay and destination characterized by double (cascade) Rayleigh fading. We successfully obtain the closed-form mathematical expressions of the outage probability (OP) and throughput of the considered energy harvesting- (EH-) FDR-V2V system. Based on these expressions, the system performance is investigated through various scenarios. Numerical results indicate that the performance of the considered system is reduced compared with that of the system over Rayleigh fading channels. We also observe that there is an optimal EH time duration that minimizes the OP and maximizes the throughput. This value depends on the transmission power of source. Furthermore, the OP goes to outage floor faster due to the impact of the residual self-interference (RSI), especially when RSI is high. All analysis results are verified by Monte-Carlo simulations.

1. Introduction

Recently, energy harvesting (EH) has attracted great attention due to the advantages and applications of this technique in wireless sensors, microcontrollers, and displays [1–4]. Compared with the traditional wireline power supply, the wireless power supply using EH can be applied for the devices whose batteries cannot be charged by traditional methods, such as the devices in body area networks. Therefore, harvesting energy from the environment is a promising approach to prolong the lifetime of energy-constrained wireless networks. Among other renewable energy sources such as solar and wind, background radio-frequency (RF) signals radiated by ambient transmitters can be a viable new source for wireless power transfer (WPT). On the other hand, RF signals have been widely used as a vehicle for wireless information transmission (WIT). Simultaneous wireless information and power transfer (SWIPT) becomes appealing since it

realizes both useful utilizations of RF signals at the same time and thus potentially offers great convenience to mobile users.

Meanwhile, full-duplex (FD) communications can increase the spectral efficiency two times compared with the traditional half-duplex (HD) communications because FD devices simultaneously transmit and receive signal at the same time and on the same frequency band. Thus, the FD technique is able to satisfy the requirement of the future wireless network such as the fifth generation (5G) and beyond [5–8]. However, the residual self-interference (RSI) after self-interference cancellation (SIC) increases the outage and bit error rate (BER) and reduces the capacity of FD communication systems. Fortunately, the researches and measurements have demonstrated that the FD systems can be deployed in realistic scenarios with the help of the recent advantage techniques such as antenna design, analog, and digital signal processing. By using three domains for cancellation such as antenna domain suppression, analog, and digital domain

cancellation, the self-interference (SI) can be suppressed up to 110 dB, making the RSI as small as the noise floor [9, 10]. Nowadays, the FD transmission is exploited at relay to enhance the coverage and reliability of the wireless communication systems [6, 11–14]. It is demonstrated that by using FD relay (FDR), the wireless systems can get higher capacity with small performance loss compared with the traditional HD relay (HDR) wireless systems.

To prolong the lifetime of wireless devices in FDR communication systems, the EH technique has been considered in many research works in the literature [3, 12, 15–19]. Based on mathematical analysis, the performance of EH-FDR systems has been investigated in various scenarios such as over Nakagami- m fading channel [12], Rayleigh fading [3, 15–18], and $\alpha - \mu$ environment [19]. These works derived the analysis expressions of the outage probability (OP) [3, 12, 15–19], symbol error probability (SEP) [12, 15], and ergodic capacity [16, 18] of the EH-FDR systems in the case that only FD relay harvests the energy from source or both source and FD relay harvest the energy from power beacon (PB). Numerical results indicated that these systems can operate in practical scenarios with high transmission power of source or PB. Moreover, using PB with multiple antennas can significantly increase the harvested energy at source and FD relay, thus improving the OP and SEP performance. On the other hand, the RSI due to FD transmission mode has a great impact on the performance of EH-FDR systems. It makes the OP and SEP of the EH-FDR systems go to error floor in high signal-to-noise ratio (SNR) regime. In addition, choosing a suitable EH time duration is very important because that value can minimize the OP and SEP of the EH-FDR systems.

Additionally, various works in the literature have proposed several architectures for SWIPT receivers such as ideal receiver, time switching (TS) receiver, and power splitting (PS) receiver [20, 21]. The advantages and disadvantages of these SWIPT receivers have been discussed. Specifically, the ideal receiver is not suitable for practical circuits, due to the fact that the receiver could not be able to decode information and harvest energy from the same signal [21]. Meanwhile, the TS and PS receivers are applicable in practical. In TS protocol, the receiver harvests energy and retrieves information in two time slots. In PS protocol, the receiver splits the received signal into two streams, one for EH and the other for information decoder [21].

Today, vehicle-to-vehicle (V2V) communication systems are fast developed due to their crucial role in the road safety applications, especially in the cooperative driving and semi-autonomous driving [22]. In that context, FDR is used to reduce the transmission delay between vehicles in intelligent transportation systems (ITS) [22–24]. In the literature, various works have investigated the performance of FDR-V2V systems via the mathematical expressions of OP and SEP [23, 24] and proposed several solutions such as antenna design [25] and interference management [26] to improve the performance of FDR-V2V systems. These works indicated that in the case of V2V communications, the channels between vehicles are not the traditional channels such as Rayleigh, Nakagami, and Rician. Instead, the double Rayleigh fading channels best describe the characteristics of V2V com-

munication. Furthermore, under the impacts of both the RSI and V2V channels, the OP and SEP of FDR-V2V systems go to the error floor faster [23, 24].

On the other hand, when wireless devices move on the road and transmit signals for a long time, traditional power supply methods, such as wireline power supply, may be impossible to be used. Therefore, EH technique becomes an effective power supply for FDR-V2V communication systems. When equipped with EH circuits, the V2V devices can move and harvest the energy at the same time. Then, they use that harvested energy to transmit signals. However, due to the computational complexity caused by the double Rayleigh fading channels, especially when EH is applied, the research about the EH-FDR-V2V system is still limited.

Motivated by these above matters, we mathematically evaluate the performance of an EH-FDR-V2V system over double Rayleigh fading channels. Although the combination of EH and FDR into a traditional wireless communication system has been investigated in various works, this is the first work that considers the combination of EH and FDR into a V2V communication system where the FD relay harvests energy from source and uses DF protocol. We successfully derive the exact expressions of the OP and throughput of the considered EH-FDR-V2V system. The main contributions of the paper can be summarized as follows:

- (1) A model of the EH-FDR-V2V system is investigated where source is a static base station while relay and destination are moving vehicles. In addition, the relay harvests the energy from source via RF signals while moving on the road; thus, the channels from relay to destination are double Rayleigh fading channels. Moreover, the relay uses decode-and-forward (DF) protocol to forward data from source to destination
- (2) The exact closed-form expressions of OP and throughput of the considered EH-FDR-V2V system under the impact of both RSI and double Rayleigh fading channels are obtained. We show that, compared with the Rayleigh fading channels, the double Rayleigh fading channels make the mathematical derivations more difficult
- (3) The performance of the considered EH-FDR-V2V system is investigated in various scenarios. Numerical results show that the considered system is greatly reduced because of the double Rayleigh fading channels. With a certain transmission power of source, there is an optimal EH time duration which minimizes the OP and maximizes the throughput of the considered system. Furthermore, higher RSI makes the OP goes to outage floor faster. Therefore, based on the transmission power of source and the system requirements, we can choose a suitable value of the EH time duration to achieve both higher OP performance and throughput. Finally, we validate all analysis results by Monte-Carlo simulations

The rest of this paper is organized as follows: Section 2 describes the system and signal model of the considered

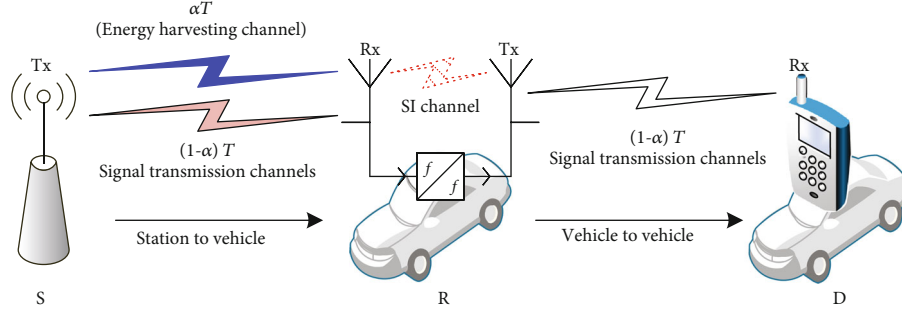


FIGURE 1: System and signal model of the considered EH-FDR-V2V system.

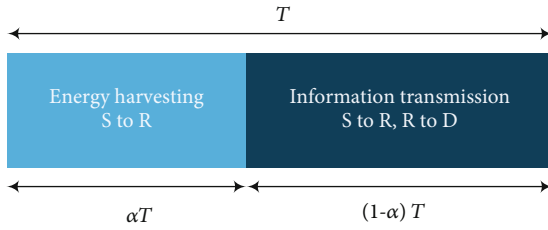


FIGURE 2: Operation of TS protocol at SWIPT receiver.

EH-FDR-V2V system. Then, Section 3 analyzes the performance of the considered system by deriving the exact closed-form expressions of OP and throughput. Section 4 presents the numerical results and discussions. Finally, Section 5 concludes the paper.

2. System Models

Figure 1 presents the system model of the considered EH-FDR-V2V system. Data is transmitted from a static source (S) to a destination (D), a moving vehicle, via the assistance of relay (R), also a moving vehicle. S and D have one antenna and operate in HD mode while R has two antennas and operates in FD mode. In fact, R can use only one antenna for both transmitting and receiving, namely, shared antenna. However, the usage of separate antennas improves SIC capability because various methods for SIC in antenna domain such as isolation, antenna directionality, and cross-polarization can be easily applied [27]. Since the power supply of R is limited, especially when R moves on the road and exchanges data for a long time, R needs to harvest energy from the RF signals transmitted by S and then uses all the harvested energy for signal transmission.

There are two EH protocols commonly used in wireless systems, i.e., TS and PS protocols as mentioned previously. In addition, the combination of these two protocols has also been applied in the literature to increase the amount of harvested energy and the performance of SWIPT systems. Various works such as [28–30] demonstrated that PS protocol can provide better secrecy outage probability (SOP) and ergodic secrecy rate than TS protocol. However, the throughput of TS protocol outperforms that of PS protocol at relatively low signal-to-noise-ratio (SNR) and high trans-

mission rate [31]. Thus, we will use TS protocol for the analysis in this paper.

On the other hand, since S is stationary and R is a moving vehicle, the channel between S and R is influenced by Rayleigh fading because there is only one group of scatterers around the receiver [32, 33]. Meanwhile, since both R and D are moving vehicles, the channel between R and D is affected by double (cascade) Rayleigh fading because there are two independent groups of scatterers around both the transmitter and receiver [32–37].

Figure 2 illustrates the operation of TS protocol at the SWIPT receiver in transmission block T . It comprises of two stages: EH and data transmission. In the first stage, the time duration of αT with time switching ratio $0 \leq \alpha \leq 1$ is used for R to harvest the energy from the RF signals transmitted from S. In this stage, S can optimize the transmitted waveforms to enhance the harvested energy at R. Meanwhile, in the second stage, the time duration of $(1 - \alpha)T$ is used for data transmission from S to R and from R to D. Since R simultaneously receives signals from S and transmits them to D at the same time and on the same frequency band, the SI from transmission antenna to reception antenna of R occurs. With separate antenna for transmission and reception, R can have higher SIC capability compared with shared antennas. Furthermore, R can use both antennas for EH in the time duration αT to obtain higher the harvested energy. However, the work in [38] demonstrated that the usage of both antennas for EH will reduce the system performance compared with the case that only one antenna is used for EH. It is because higher harvested energy leads to higher transmission power of R and thus higher SI power. Consequently, in this paper, we consider the case that only one antenna of R is used for EH.

In the time duration αT for EH, the harvested energy at R (denoted by E_h^R) is given by [1]

$$E_h^R = \frac{\eta \alpha T P_S |h_{SR}|^2}{d_{SR}^m}, \quad (1)$$

where P_S is the average transmission power of S; h_{SR} is the fading coefficient of channel from S to R; d_{SR} is the distance between S and R; $2 \leq m \leq 6$ is the path loss exponent; and $0 \leq \eta \leq 1$ is the energy conversion efficiency and its value is a constant but depends on the quality of electronic circuit.

Because all the harvested energy is used for data transmission, the transmission power of R is computed as

$$P_R = \frac{\eta\alpha TP_S |h_{SR}|^2}{d_{SR}^m (1-\alpha) T} = \frac{\eta\alpha P_S |h_{SR}|^2}{d_{SR}^m (1-\alpha)}. \quad (2)$$

In practice, the output power of EH circuit may be proportional to the harvested input power up to a certain saturation power threshold P_{th} . When the input power exceeds P_{th} , the output power remains unchanged. There are various factors causing the nonlinear characteristic of an energy harvester such as diode and saturation nonlinearities. Nonlinearities are an intrinsic property of diode. Meanwhile, saturation nonlinearities were confirmed by experiments on energy harvesting circuits [21, 39–41]. As a result of this nonlinear characteristic, the transmission power of R with a nonlinear energy harvester can be obtained by extending (2) as

$$P_R = \begin{cases} \frac{\eta\alpha P_S |h_{SR}|^2}{d_{SR}^m (1-\alpha)}, & \frac{P_S |h_{SR}|^2}{d_{SR}^m} \leq P_{th}, \\ \frac{\eta\alpha P_{th}}{1-\alpha}, & \frac{P_S |h_{SR}|^2}{d_{SR}^m} > P_{th}. \end{cases} \quad (3)$$

In the time duration $(1-\alpha)T$, R receives signals from S while it forwards signals to D. This operation creates SI from the transmission antenna to the reception antenna of R. The received signals at R and D are now expressed as

$$y_R = \frac{1}{\sqrt{d_{SR}^m}} h_{SR} \sqrt{P_S} x_S + \frac{1}{\sqrt{d_{RR}^m}} \tilde{h}_{RR} \sqrt{P_R} x_R + z_R, \quad (4)$$

$$y_D = \frac{1}{\sqrt{d_{RD}^m}} h_{RD} \sqrt{P_R} x_R + z_D, \quad (5)$$

where d_{RR} and d_{RD} are, respectively, the distances from the transmission to reception antennas of R and from R to D; h_{SR} , h_{RD} , and \tilde{h}_{RR} are, respectively, the fading coefficients of S-R channel and R-D channel and from the transmission antenna to the reception antenna of R; x_S and x_R are the transmitted signals at S and R, respectively; and P_S and P_R are, respectively, the average transmission powers of S and R; z_R and z_D are the Gaussian noises with zero mean and variance of σ^2 , i.e., $z_R \sim \mathcal{CN}(0, \sigma^2)$ and $z_D \sim \mathcal{CN}(0, \sigma^2)$.

As shown in (4), the term $(1/\sqrt{d_{RR}^m})\tilde{h}_{RR}\sqrt{P_R}x_R$ is the SI. The average power of SI before SIC can be computed as

$$E \left\{ \frac{|\tilde{h}_{RR}|^2 P_R}{d_{RR}^m} \right\} = \frac{\eta\alpha P_S}{1-\alpha} E \left\{ \frac{|\tilde{h}_{RR}|^2 |h_{SR}|^2}{d_{RR}^m} \right\}, \quad (6)$$

where $E\{\cdot\}$ is the expectation operator.

Since R can apply all SIC techniques in three domains such as antenna propagation, analog suppression, and digital cancellation, the SI power is greatly reduced. Through SI channel estimation, SI can be suppressed effectively, especially in the digital cancellation domain. However, due to

imperfect SIC, the residual SI (RSI) still exists at R. According to the analysis and measurements, the RSI due to FD mode (denoted by I_{RSI}) is modeled by a complex Gaussian distribution with zero mean and variance of σ_{RSI}^2 [3, 8, 9, 42, 43], where σ_{RSI}^2 is given by

$$\sigma_{RSI}^2 = \frac{k\eta\alpha P_S}{1-\alpha}, \quad (7)$$

where k denotes the SIC capability of the FD relay.

After all SIC techniques, (4) can be rewritten as

$$y_R = \frac{1}{\sqrt{d_{SR}^m}} h_{SR} \sqrt{P_S} x_S + I_R + z_R. \quad (8)$$

Based on (8) and (5), the signal-to-interference-plus-noise ratios (SINRs) at R (denoted by γ_R) and D (denoted by γ_D) of the considered EH-FDR-V2V system are, respectively, computed as

$$\gamma_R = \frac{|h_{SR}|^2 P_S}{d_{SR}^m (\sigma_{RSI}^2 + \sigma^2)}, \quad (9)$$

$$\gamma_D = \frac{|h_{RD}|^2 P_R}{d_{RD}^m \sigma^2} = \frac{|h_{SR}|^2 |h_{RD}|^2 \eta\alpha P_S}{d_{SR}^m d_{RD}^m \sigma^2 (1-\alpha)}.$$

When DF protocol is applied at FD relay, the end-to-end SINR (denoted by γ) of the considered system is calculated as

$$\gamma = \min(\gamma_R, \gamma_D). \quad (10)$$

3. System Performance Analysis

3.1. Outage Probability Analysis. In this subsection, we derive the analysis expression of the OP for evaluating the system performance. Mathematically, the OP is calculated as

$$\begin{aligned} \text{OP} &= \Pr \{ (1-\alpha) \log_2(1+\gamma) < \mathcal{R} \} \\ &= \Pr \left\{ \gamma < 2^{\mathcal{R}/(1-\alpha)} - 1 \right\} \\ &= \Pr \{ \gamma < \gamma_{th} \}, \end{aligned} \quad (11)$$

where \mathcal{R} (bit/s/Hz) is a predata transmission rate; γ is the end-to-end SINR of the considered EH-FDR-V2V system given in (10); and $\gamma_{th} = 2^{\mathcal{R}/(1-\alpha)} - 1$ is the SINR threshold.

From (11), the OP of the considered EH-FDR-V2V system is derived in Theorem 1.

Theorem 1. *Under the impact of the RSI, the OP of the considered EH-FDR-V2V system over double Rayleigh fading channels is given by*

$$OP = 1 - \frac{\pi}{2N} \exp\left(-\frac{\Lambda\gamma_{th}}{\Omega_1}\right) \sum_{n=1}^N \sqrt{\frac{\Theta\gamma_{th}(1-\phi_n^2)}{\Omega_2\Omega_3(\Omega_1 \ln(1/u) + \Lambda\gamma_{th})}} \cdot K_1\left(\sqrt{\frac{\Theta\gamma_{th}}{\Omega_2\Omega_3(\Omega_1 \ln(1/u) + \Lambda\gamma_{th})}}\right), \quad (12)$$

where $\Lambda = (d_{SR}^m(\sigma_{RSI}^2 + \sigma^2))/P_S$; $\Theta = (4d_{SR}^m d_{RD}^m \sigma^2(1-\alpha))/\eta\alpha P_S$; Ω_1 is the average channel gain of Rayleigh fading S-R channel; Ω_2 and Ω_3 are the average channel gains of the double Rayleigh R-D channels; N is the complexity-accuracy trade-off parameter; $\phi_n = \cos((2n-1)\pi/2N)$; $u = (1/2)$

$(\phi_n + 1)$; and $K_1(\cdot)$ denotes the first-order modified Bessel function of the second kind [44].

Proof. Replacing γ in (10) into (11), we have

$$OP = \Pr\{\min(\gamma_R, \gamma_D) < \gamma_{th}\} = 1 - \Pr\{\gamma_R > \gamma_{th}, \gamma_D > \gamma_{th}\}. \quad (13)$$

Thus, it is obvious that the OP occurs when S-R or R-D link is in outage because R is out of the coverage area of S or D is out of the coverage area of R.

Substituting γ_R and γ_D in (9) into (13), we rewrite (13) as

$$\begin{aligned} OP &= 1 - \Pr\left\{\frac{|h_{SR}|^2 P_S}{d_{SR}^m(\sigma_{RSI}^2 + \sigma^2)} > \gamma_{th}, \frac{|h_{SR}|^2 |h_{RD}|^2 \eta\alpha P_S}{d_{SR}^m d_{RD}^m \sigma^2(1-\alpha)} > \gamma_{th}\right\} \\ &= 1 - \Pr\left\{|h_{SR}|^2 > \frac{\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2)}{P_S}, |h_{RD}|^2 > \frac{\gamma_{th} d_{SR}^m d_{RD}^m \sigma^2(1-\alpha)}{|h_{SR}|^2 \eta\alpha P_S}\right\} \\ &= 1 - \left(1 - \Pr\left\{|h_{RD}|^2 \leq \frac{\gamma_{th} d_{SR}^m d_{RD}^m \sigma^2(1-\alpha)}{|h_{SR}|^2 \eta\alpha P_S} \mid |h_{SR}|^2 > \left(\frac{\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2)}{P_S}\right)\right\}\right). \end{aligned} \quad (14)$$

Applying the property of conditional probability [45], (14) becomes

$$OP = 1 - \int_0^\infty \left(1 - F_{|h_{RD}|^2}\left(\frac{\gamma_{th} d_{SR}^m d_{RD}^m \sigma^2(1-\alpha)}{|h_{SR}|^2 \eta\alpha P_S}\right)\right) \cdot f_{|h_{SR}|^2}\left(y + \frac{\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2)}{P_S}\right) dy, \quad (15)$$

where $y = |h_{SR}|^2 - ((\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2))/P_S)$.

For calculating the integral in (15), we need to derive the distributions of $|h_{SR}|^2$ and $|h_{RD}|^2$. Since S-R channel is influenced by Rayleigh fading, the cumulative distribution function (CDF, denoted by $F(\cdot)$) and the probability density function (PDF, denoted by $f(\cdot)$) of $|h_{SR}|^2$ are given by

$$F_{|h_{SR}|^2}(x) = 1 - \exp\left(-\frac{x}{\Omega_1}\right), \quad x \geq 0, \quad (16)$$

$$f_{|h_{SR}|^2}(x) = \frac{1}{\Omega_1} \exp\left(-\frac{x}{\Omega_1}\right), \quad x \geq 0, \quad (17)$$

where $\Omega_1 = E\{|h_{SR}|^2\}$ is the average channel gain of S-R communication link.

Furthermore, since R-D channel is double Rayleigh fading channel, thus, $|h_{RD}|^2$ is considered the multiplication of two independent variables $|h_2|^2$ and $|h_3|^2$, which means

$|h_{RD}|^2 = |h_2|^2 |h_3|^2$ where $|h_2|^2$ and $|h_3|^2$ are the instantaneous channel gains of the Rayleigh fading channel with the average channel gains $\Omega_2 = E\{|h_2|^2\}$ and $\Omega_3 = E\{|h_3|^2\}$. The CDF and PDF of $|h_{RD}|^2$ are, respectively, given by [23, 24, 46]

$$F_{|h_{RD}|^2}(x) = 1 - \sqrt{\frac{4x}{\Omega_2\Omega_3}} K_1\left(\sqrt{\frac{4x}{\Omega_2\Omega_3}}\right), \quad (18)$$

$$f_{|h_{RD}|^2}(x) = \frac{2}{\Omega_2\Omega_3} K_0\left(\sqrt{\frac{4x}{\Omega_2\Omega_3}}\right), \quad (19)$$

where $K_0(\cdot)$ is the zero-order modified Bessel function of the second kind [44].

Now, applying (16), (17), (18), and (19) to calculate (15), we have

$$\begin{aligned} OP &= 1 - \int_0^\infty \sqrt{\frac{4\gamma_{th} d_{SR}^m d_{RD}^m \sigma^2(1-\alpha)}{\Omega_2\Omega_3(y + ((\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2))/P_S))\eta\alpha P_S}} K_1 \\ &\quad \cdot \left(\sqrt{\frac{4\gamma_{th} d_{SR}^m d_{RD}^m \sigma^2(1-\alpha)}{\Omega_2\Omega_3(y + ((\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2))/P_S))\eta\alpha P_S}}\right) \\ &\quad \times \frac{1}{\Omega_1} \exp\left(-\frac{y}{\Omega_1} - \frac{\gamma_{th} d_{SR}^m(\sigma_{RSI}^2 + \sigma^2)}{\Omega_1 P_S}\right) dy \end{aligned}$$

$$\begin{aligned}
&= 1 - \frac{1}{\Omega_1} \int_0^\infty \sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (y + \Lambda \gamma_{\text{th}})}} K_1 \\
&\quad \cdot \left(\sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (y + \Lambda \gamma_{\text{th}})}} \right) \exp\left(-\frac{y}{\Omega_1} - \frac{\Lambda \gamma_{\text{th}}}{\Omega_1}\right) dy \quad (20) \\
&= 1 - \frac{1}{\Omega_1} \exp\left(-\frac{\Lambda \gamma_{\text{th}}}{\Omega_1}\right) \int_0^\infty \sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (y + \Lambda \gamma_{\text{th}})}} K_1 \\
&\quad \cdot \left(\sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (y + \Lambda \gamma_{\text{th}})}} \right) \exp\left(-\frac{y}{\Omega_1}\right) dy.
\end{aligned}$$

By changing the variable in (20), i.e., $z = \exp(-y/\Omega_1)$, (20) now becomes

$$\begin{aligned}
\text{OP} &= 1 - \exp\left(-\frac{\Lambda \gamma_{\text{th}}}{\Omega_1}\right) \int_0^1 \sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (\Omega_1 \ln(1/z) + \Lambda \gamma_{\text{th}})}} K_1 \\
&\quad \cdot \left(\sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (\Omega_1 \ln(1/z) + \Lambda \gamma_{\text{th}})}} \right) dz. \quad (21)
\end{aligned}$$

Using the Gaussian-Chebyshev quadrature method [47] for calculating the integral in (21), we have

$$\begin{aligned}
&\int_0^1 \sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (\Omega_1 \ln(1/z) + \Lambda \gamma_{\text{th}})}} K_1 \\
&\quad \cdot \left(\sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (\Omega_1 \ln(1/z) + \Lambda \gamma_{\text{th}})}} \right) dz \quad (22) \\
&= \frac{\pi}{2N} \sum_{n=1}^N \sqrt{\frac{\Theta \gamma_{\text{th}} (1 - \phi_n^2)}{\Omega_2 \Omega_3 (\Omega_1 \ln(1/u) + \Lambda \gamma_{\text{th}})}} K_1 \\
&\quad \cdot \left(\sqrt{\frac{\Theta \gamma_{\text{th}}}{\Omega_2 \Omega_3 (\Omega_1 \ln(1/u) + \Lambda \gamma_{\text{th}})}} \right),
\end{aligned}$$

where N , ϕ_n , and u are defined after (12).

Plugging (22) into (21), we obtain the OP of the considered system as in (12). The proof is complete.

3.2. Throughput Analysis. For a wireless system, besides the OP, throughput is another important parameter that needs to be evaluated. Therefore, in this subsection, we derive the throughput of the considered EH-FDR-V2V system.

Theorem 2. *The throughput (denoted by \mathcal{T}_{put}) of the considered EH-FDR-V2V system is calculated as*

$$\mathcal{T}_{\text{put}} = \mathcal{R}(1 - \alpha)(1 - \text{OP}), \quad (23)$$

where \mathcal{R} and α are, respectively, the predata transmission rate and the time switching ratio and OP is the outage probability, which is given in (12).

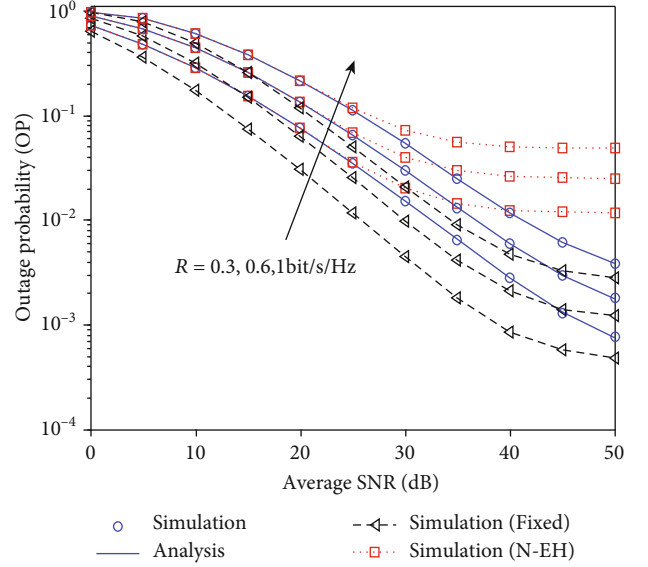


FIGURE 3: The impact of double Rayleigh fading channels on the OP of the considered EH-FDR-V2V system with $\mathcal{R} = 0.3, 0.6$, and 1 bit/s/Hz , $k = -30 \text{ dB}$, and $\alpha = 0.5$.

4. Numerical Results and Discussion

In this section, we use the OP and throughput expressions obtained in the previous section to evaluate the performance of the considered EH-FDR-V2V system. Various scenarios are carried out to investigate the impacts of the double Rayleigh fading channels, the RSI, and the time switching ratio on the system performance. The Monte-Carlo simulations are used to verify the correctness of our analysis. In all results, we set the average SNR $= P_S/\sigma^2$, the energy harvesting efficiency is $\eta = 0.85$, the path loss exponent is $m = 2.7$ (that is similar for an urban cellular network environment [31]), and the distances (d_{SR} and d_{RD}) and the average channel gains (Ω_1 , Ω_2 , and Ω_3) are normalized to a unit value such as in [14, 31, 42].

Figure 3 plots the OP of the considered EH-FDR-V2V system versus the average SNR using (12) in Theorem 1. To evaluate the impact of double Rayleigh fading channels, we also provide the OP of this system in the case that R and D are static (denoted by “Fixed” in Figure 3). Note that in this case, S-R and R-D channels are influenced by Rayleigh fading. If S and D in [17] were equipped with a single antenna, the system model in [17] would become the case of “Simulation (Fixed)” in this paper. Moreover, we simulate the OP of the considered EH-FDR-V2V system with a nonlinear energy harvester and the saturation power threshold is $P_{\text{th}}/\sigma^2 = 25 \text{ dB}$ to compared with the OP of this system with a linear energy harvester. As can be seen from Figure 3, with $\mathcal{R} = 0.3 \text{ bit/s/Hz}$ and $\text{OP} = 10^{-3}$, the OP performance of the considered EH-FDR-V2V system over double Rayleigh fading channels is 5 dB lower compared with that over Rayleigh fading channels. With a higher data transmission rate, e.g., $\mathcal{R} = 0.6$ and $\mathcal{R} = 1 \text{ bit/s/Hz}$, the OP in the case of double Rayleigh fading channels still goes down while the OP in the case of Rayleigh fading channels nearly goes to outage

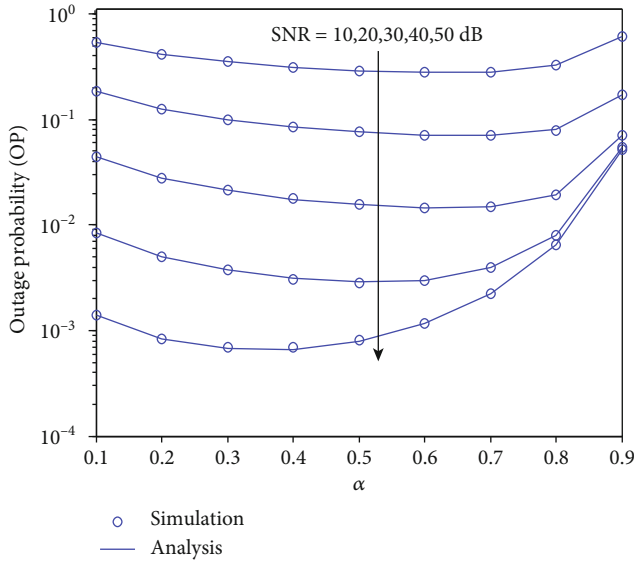


FIGURE 4: The OP of the considered EH-FDR-V2V system versus the time switching ratio α with different values of SNR, $k = -30$ dB, and $\mathcal{R} = 0.3$ bit/s/Hz.

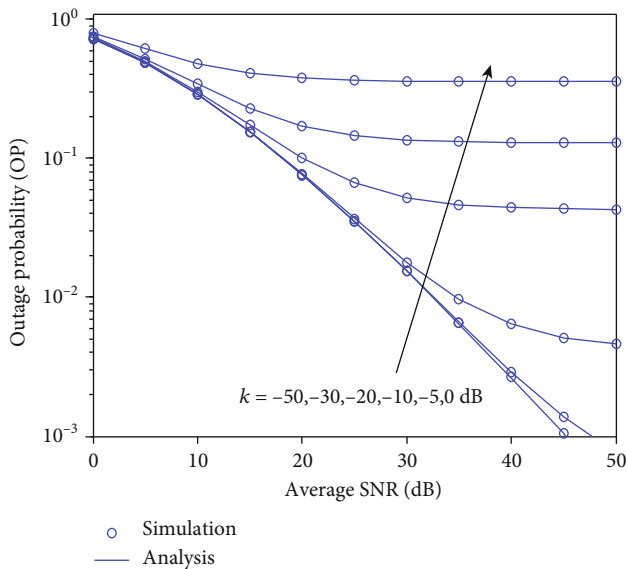


FIGURE 5: The impact of SIC capability on the OP of the considered EH-FDR-V2V system with $\mathcal{R} = 0.3$ bit/s/Hz and $\alpha = 0.5$.

floor due to the impact of the RSI. Furthermore, due to the effect of saturation power threshold, the OPs with nonlinear energy harvesters go to the floors faster than the OPs with linear energy harvesters as $\text{SNR} \geq 35$ dB.

Figure 4 shows the OP of the considered EH-FDR-V2V system versus the time switching ratio α with different values of SNR, i.e., $\text{SNR} = 10, 20, 30, 40,$ and 50 dB. For a certain value of SNR, there is an optimal α that minimizes the OP of the considered system. For example, in the case of $\text{SNR} = 50$ dB, the optimal $\alpha = 0.3$. When SNR decreases, the optimal α is increased, i.e., $\alpha = 0.5$ for $\text{SNR} = 40$ dB, $\alpha = 0.6$ for $\text{SNR} = 30$ dB, and $\alpha = 0.7$ for $\text{SNR} = 20, 10$ dB. These results

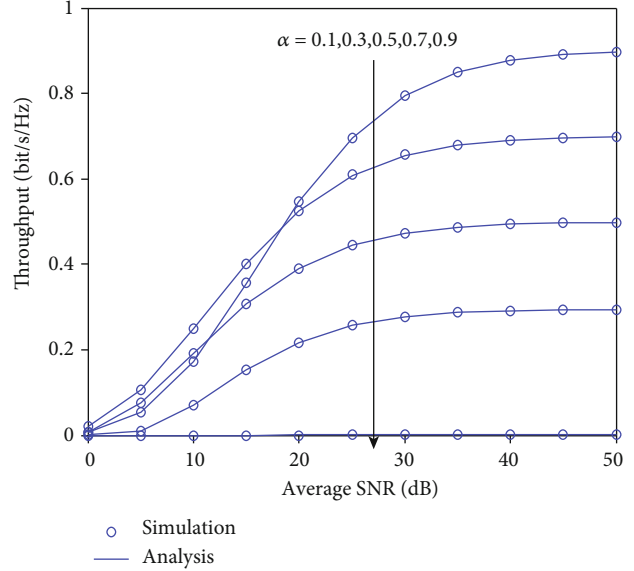


FIGURE 6: The throughput of the considered EH-FDR-V2V system versus the average SNR with different values of time switching ratio, $\mathcal{R} = 1$ bit/s/Hz, and $k = -30$ dB.

are reasonable for the considered system because it is obvious that, in low SNR regime (low transmission power of S), R needs a long time to harvest enough energy for transmitting signals. In the case of high SNR regime, R can harvest enough energy in a short time; thus, the optimal point is reduced. Based on these features, we can choose a suitable value of the time switching ratio to get the lowest OP of the considered system depending on the transmission power of S.

Figure 5 investigates the impact of SIC capability k on the OP of the considered EH-FDR-V2V system. We can see in Figure 5 that with small values of k , e.g., $k = -50$ and -30 dB, the impact of the RSI on the OP is small. Therefore, the OP still goes down in the evaluated range of the SNR. However, with a higher value of k , e.g., $k = -20$ dB, the impact of the RSI is very strong, especially in the high SNR regime. Particularly, in the range from 0 to 30 dB of the SNR, the OPs with $k = -50, -30,$ and -20 dB are very similar. However, when $\text{SNR} > 35$ dB, the differences between the OPs with $k = -50$ dB, $k = -30$ dB, and $k = -20$ dB are remarkable. For $k = -20$ dB, the OP goes to outage floor when $\text{SNR} > 50$ dB. For a higher value of k , e.g., $k = -10, -5,$ and 0 dB, the OPs go to outage floor earlier (at $\text{SNR} = 35$ dB). The outage floors are $4 \times 10^{-2}, 1.3 \times 10^{-1},$ and 3.6×10^{-1} corresponding to $k = -10$ dB, $k = -5$ dB, and $k = 0$ dB, respectively. Therefore, all SIC solutions must be applied effectively to get the smallest RSI.

Figure 6 illustrates the throughput of the considered EH-FDR-V2V system versus the average SNR for various time switching ratio α . We use (23) in Theorem 2 to plot the analysis curves of the throughput. We show that lower α leads to higher throughput. It is because lower α means the duration time for EH is reduced; thus, the duration time for data exchange is increased. As a result, the considered system has higher throughput. With higher α , the time duration for data exchange will be reduced, leading to a reduction in

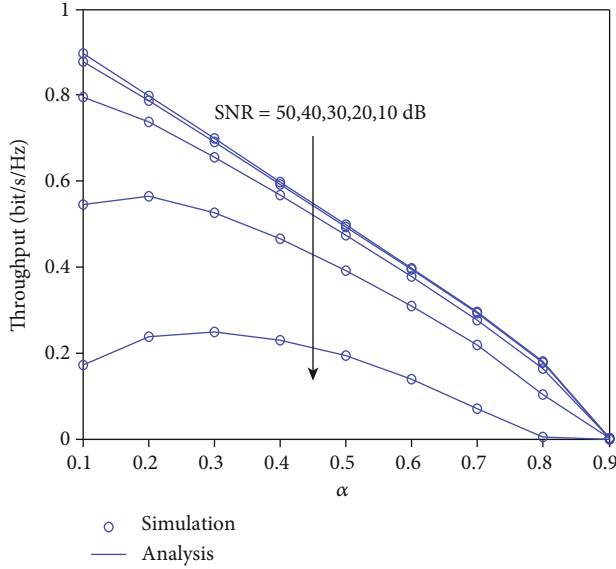


FIGURE 7: The throughput of the considered EH-FDR-V2V system versus the time switching ratio α with different values of SNR and $k = -30$ dB.

the throughput. It is also noted that, although we have higher throughput with lower α , the usage of a small value of α can reduce the system performance because the destination (D) cannot detect its messages successfully. Therefore, we need to combine both the OP performance and the throughput to select a suitable α . For the convenience in choosing α for the considered system, we will investigate the relationship between throughput and α in the next scenario.

Figure 7 investigates the throughput of the considered EH-FDR-V2V system versus the time switching ratio α . We can see that, with high SNRs, the throughput is highest when the time switching ratio α is smallest. For example, in the case of SNR = 50, 40, and 30 dB, the throughput is maximal when $\alpha = 0.1$. However, with lower SNRs, e.g., SNR = 20 and 10 dB, the value $\alpha = 0.1$ is not the optimal value that maximizes the throughput. In these cases, the throughput is highest when $\alpha = 0.2$ and $\alpha = 0.3$ for SNR = 20 dB and SNR = 10 dB, respectively. We should also remind that $\alpha = 0$ should not be used to get the maximal throughput because when $\alpha = 0$ (no time for EH), the transmission power of R also equals zero. In this case, we have $OP = 1$ and $\mathcal{F}_{\text{put}} = 0$ (refer to (23)). Based on Figures 4 and 7, depending on the requirements of OP and throughput, we can choose a suitable value of α for the considered EH-FDR-V2V system. For example, in the case of SNR = 50 dB, we can choose the range of α from 0.1 to 0.3 to get high OP performance and throughput.

5. Conclusions

Motivated by the big advantages of the EH technique, in this paper, we investigated the performance of the EH-FDR-V2V system under the impact of RSI due to the FD transmission mode over double Rayleigh fading channels. We successfully derived the closed-form expressions of the outage probability and throughput of the considered system. Based on these

expressions, the system performance is investigated through various scenarios. Numerical results showed that the performance of the considered system is greatly reduced compared with that system over Rayleigh fading channels. With a certain value of the transmission power of source, there is an optimal EH time duration that minimizes the OP and maximizes the throughput. Therefore, based on the transmission power of source and the system requirements, we can choose a suitable EH time duration to get high OP performance and throughput. Furthermore, the RSI causes the outage floor in the considered system, especially in the case of high RSI. Thus, effective SIC techniques should be used when deploying the FD systems in practice.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with RF energy harvesting: a contemporary survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 757–789, 2015.
- [2] V.-D. Nguyen, T. Q. Duong, H. D. Tuan, O.-S. Shin, and H. V. Poor, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2220–2233, 2017.
- [3] B. C. Nguyen, T. M. Hoang, P. T. Tran, and T. N. Nguyen, "Outage probability of NOMA system with wireless power transfer at source and full-duplex relay," *AEU-International Journal of Electronics and Communications*, vol. 116, article 152957, 2020.
- [4] H. H. M. Tam, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "MIMO energy harvesting in full-duplex multi-user networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3282–3297, 2017.
- [5] Y. Deng, K. J. Kim, T. Q. Duong, M. ElKashlan, G. K. Karagiannis, and A. Nallanathan, "Full-duplex spectrum sharing in cooperative single carrier systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, no. 1, pp. 68–82, 2016.
- [6] X.-T. Doan, N.-P. Nguyen, C. Yin, D. B. Da Costa, and T. Q. Duong, "Cognitive full-duplex relay networks under the peak interference power constraint of multiple primary users," *EURASIP Journal on Wireless Communications and Networking*, vol. 2017, no. 1, Article ID 8, 2017.
- [7] B. C. Nguyen, X. N. Tran, and D. T. Tran, "Performance analysis of in-band full-duplex amplify-and-forward relay system with direct link," in *2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom)*, pp. 192–197, Ho Chi Minh City, Vietnam, January 2018.
- [8] B. C. Nguyen, X. N. Tran, D. T. Tran, and L. T. Dung, "Full-duplex amplify-and-forward relay system with direct link:

- performance analysis and optimization,” *Physical Communication*, vol. 37, article 100888, 2019.
- [9] D. Bharadia, E. McMillin, and S. Katti, “Full duplex radios,” *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 375–386, 2013.
- [10] A. H. Gazestani, S. A. Ghorashi, B. Mousavinasab, and M. Shikh-Bahaei, “A survey on implementation and applications of full duplex wireless communications,” *Physical Communication*, vol. 34, pp. 121–134, 2019.
- [11] B. C. Nguyen and X. N. Tran, “Performance analysis of full-duplex amplify-and-forward relay system with hardware impairments and imperfect self-interference cancellation,” *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 4946298, 10 pages, 2019.
- [12] B. C. Nguyen, T. M. Hoang, and P. T. Tran, “Performance analysis of full-duplex decode-and-forward relay system with energy harvesting over Nakagami- m fading channels,” *AEU-International Journal of Electronics and Communications*, vol. 98, pp. 114–122, 2019.
- [13] I. Krikidis, H. A. Suraweera, P. J. Smith, and C. Yuen, “Full-duplex relay selection for amplify-and-forward cooperative networks,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4381–4393, 2012.
- [14] B. C. Nguyen, N. N. Thang, X. N. Tran, and L. T. Dung, “Impacts of imperfect channel state information, transceiver hardware, and self-interference cancellation on the performance of full-duplex mimo relay system,” *Sensors*, vol. 20, no. 6, article 1671, 2020.
- [15] B. C. Nguyen, T. M. Hoang, S.-G. Choi, and L. T. Dung, “Full-duplex relay system with energy harvesting: outage and symbol error probabilities,” in *2018 International Conference on Advanced Technologies for Communications (ATC)*, pp. 360–365, Ho Chi Minh City, Vietnam, October 2018.
- [16] Y. Alsaba, C. Y. Leow, and S. K. A. Rahim, “Full-duplex cooperative non-orthogonal multiple access with beamforming and energy harvesting,” *IEEE Access*, vol. 6, pp. 19726–19738, 2018.
- [17] T. M. Hoang, N. T. Tan, N. B. Cao, and L. T. Dung, “Outage probability of MIMO relaying full-duplex system with wireless information and power transfer,” in *2017 Conference on Information and Communication Technology (CICT)*, pp. 1–6, Gwalior, India, November 2017.
- [18] B. C. Nguyen, T. M. Hoang, X. N. Pham, and P. T. Tran, “Performance analysis of energy harvesting-based full-duplex decode-and-forward vehicle-to-vehicle relay networks with nonorthogonal multiple access,” *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 6097686, 11 pages, 2019.
- [19] G. Nauryzbayev, M. Abdallah, and K. M. Rabie, “Outage probability of the EH-based full-duplex AF and DF relaying systems in $\alpha - \mu$ environment,” in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pp. 1–6, Chicago, IL, USA, August 2018.
- [20] X. Zhou, R. Zhang, and C. K. Ho, “Wireless information and power transfer: architecture design and rate-energy tradeoff,” *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, 2013.
- [21] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, “Fundamentals of wireless information and power transfer: from RF energy harvester models to signal and system designs,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 4–33, 2019.
- [22] C. Campolo, A. Molinaro, A. O. Berthet, and A. Vinel, “Full-duplex radios for vehicular communications,” *IEEE Communications Magazine*, vol. 55, no. 6, pp. 182–189, 2017.
- [23] B. C. Nguyen, X. N. Tran, T. M. Hoang, and L. T. Dung, “Performance analysis of full-duplex vehicle-to-vehicle relay system over double-rayleigh fading channels,” *Mobile Networks and Applications*, vol. 25, no. 1, pp. 363–372, 2020.
- [24] B. C. Nguyen, T. M. Hoang, and L. T. Dung, “Performance analysis of vehicle-to-vehicle communication with full-duplex amplify-and-forward relay over double-rayleigh fading channels,” *Vehicular Communications*, vol. 19, p. 100166, 2019.
- [25] C.-X. Mao, S. Gao, and Y. Wang, “Dual-band full-duplex Tx/Rx antennas for vehicular communications,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4059–4070, 2018.
- [26] M. Yang, S.-W. Jeon, and D. K. Kim, “Interference management for in-band full-duplex vehicular access networks,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1820–1824, 2018.
- [27] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, “In-band full-duplex wireless: challenges and opportunities,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1637–1652, 2014.
- [28] S. Sharma, S. D. Roy, and S. Kundu, “Secrecy outage in a two-hop decode and forward relay network with accumulated harvested energy,” *Physical Communication*, vol. 36, article 100792, 2019.
- [29] S. Sharma, S. D. Roy, and S. Kundu, “Secure communication with energy harvesting multiple half-duplex DF relays assisted with jamming,” *Wireless Networks*, vol. 26, pp. 1151–1164, 2020.
- [30] S. Sharma, S. Dhar Roy, and S. Kundu, “Secure communication in cognitive radio networks with untrusted AF relays,” *International Journal of Communication Systems*, vol. 32, no. 7, article e3919, 2019.
- [31] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, “Relaying protocols for wireless energy harvesting and information processing,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3622–3636, 2013.
- [32] I. Z. Kovacs, P. C. F. Eggers, K. Olesen, and L. G. Petersen, “Investigations of outdoor-to-indoor mobile-to-mobile radio communication channels,” in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 1, pp. 430–434, Vancouver, BC, Canada, 2002.
- [33] J. Salo, H. M. El-Sallabi, and P. Vainikainen, “Statistical analysis of the multiple scattering radio channel,” *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3114–3124, 2006.
- [34] J. Salo, H. M. El-Sallabi, and P. Vainikainen, “Impact of double-Rayleigh fading on system performance,” in *2006 1st International Symposium on Wireless Pervasive Computing*, pp. 1–5, Phuket, Thailand, 2006.
- [35] T. T. Duy, G. C. Alexandropoulos, V. T. Tung, V. N. Son, and T. Q. Duong, “Outage performance of cognitive cooperative networks with relay selection over double-Rayleigh fading channels,” *IET Communications*, vol. 10, no. 1, pp. 57–64, 2016.
- [36] A. S. Akki and F. Haber, “A statistical model of mobile-to-mobile land communication channel,” *IEEE Transactions on Vehicular Technology*, vol. 35, no. 1, pp. 2–7, 1986.

- [37] I. Kovacs, *Radio Channel Characterisation for Private Mobile Radio Systems: Mobile-to-Mobile Radio Link Investigations*, [Ph.D. Thesis], Aalborg Universitet, 2002.
- [38] C. Zhong, H. A. Suraweera, G. Zheng, I. Krikidis, and Z. Zhang, "Wireless information and power transfer with full duplex relaying," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3447–3461, 2014.
- [39] Y. Dong, M. J. Hossain, and J. Cheng, "Performance of wireless powered amplify and forward relaying over Nakagami- fading channels with nonlinear energy harvester," *IEEE Communications Letters*, vol. 20, no. 4, pp. 672–675, 2016.
- [40] Z. Wei, S. Sun, X. Zhu, D. I. Kim, and D. W. K. Ng, "Resource allocation for wireless-powered full-duplex relaying systems with nonlinear energy harvesting efficiency," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12079–12093, 2019.
- [41] K. Xu, M. Zhang, J. Liu, N. Sha, W. Xie, and L. Chen, "SWIPT in mMIMO system with non-linear energy-harvesting terminals: protocol design and performance optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, Article ID 72, 2019.
- [42] C. Li, Z. Chen, Y. Wang, Y. Yao, and B. Xia, "Outage analysis of the full-duplex decode-and-forward two-way relay system," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4073–4086, 2017.
- [43] X. N. Tran, B. C. Nguyen, and D. T. Tran, "Outage probability of two-way full-duplex relay system with hardware impairments," in *2019 3rd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom)*, pp. 135–139, Hanoi, Vietnam, 2019.
- [44] A. Jeffrey and D. Zwillinger, *Table of Integrals, Series, and Products*, Academic Press, 2007.
- [45] A. Leon-Garcia and A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, Pearson/Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 2008.
- [46] Y. Ai, M. Cheffena, A. Mathur, and H. Lei, "On Physical Layer Security of Double Rayleigh Fading Channels for Vehicular Communications," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 1038–1041, 2018.
- [47] F. B. Hildebrand, *Introduction to Numerical Analysis*, Courier Corporation, 1987.