Wireless Communications and Mobile Computing

Smart Antennas and Intelligent Sensors Based Systems: Enabling Technologies and Applications

Lead Guest Editor: Fawad Zaman
Guest Editors: Sungchang Lee, Mohamad K. A. Rahim, and Sarmadullah Khan

Smart Antennas and Intelligent Sensors Based Systems: Enabling Technologies and Applications

Wireless Communications and Mobile Computing

Smart Antennas and Intelligent Sensors Based Systems: Enabling Technologies and Applications

Lead Guest Editor: Fawad Zaman
Guest Editors: Sungchang Lee, Mohamad K. A. Rahim, and Sarmadullah Khan

Editorial Board

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 6475832, 3 pages
https://doi.org/10.1155/2019/6475832

Editorial

Smart Antennas and Intelligent Sensors Based Systems: Enabling Technologies and Applications

Fawad Zaman,[1] **Sungchang Lee**,[2] **Mohamad K. A. Rahim**,[3] **and Sarmadullah Khan**[4]

[1]*Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Attock, Pakistan*
[2]*Korea Aerospace University, Goyang-si, Republic of Korea*
[3]*Universiti Teknologi Malaysia (UTM), Johor, Malaysia*
[4]*De Montfort University, Leicester, UK*

Correspondence should be addressed to Fawad Zaman; fawad@ciit-attock.edu.pk

Received 3 June 2019; Accepted 3 June 2019; Published 14 July 2019

The growing communication and computing capabilities in the devices enlarge the connected world and improve the human life comfort level. The evolution of intelligent sensor networks and smart antennas has led to the development of smart devices and systems for real-time monitoring of various environments. The demand of smart antennas and intelligent sensors significantly increases when dealing with multiuser communication system that needs to be adaptive, especially in unknown adverse environment [1–3]. The smart antennas based arrays are capable of steering the main beam in any desired direction while placing nulls in the unwanted directions. Intelligent sensor networks integration with smart antennas will provide algorithms and interesting application to collect various data of environment to make intelligent decisions [4, 5].

The aim of this special issue is to provide an inclusive vision on the current research in the area of intelligent sensors and smart antenna based systems for enabling various applications and technologies. We cordially invite some researchers to contribute papers that discuss the issues arising in intelligent sensors and smart antenna based system. Hence, this special issue offers the state-of-the-art research in this field.

The paper titled "Compressive Sensing Based Channel Estimation for Massive MIMO Communication Systems" discusses the advantage of spatial and temporal common sparsity of massive MIMO channels in delay domain, nonorthogonal pilot design, and channel estimation schemes which are proposed under the frame work of structured compressive sensing (SCS) theory that considerably reduces the pilot overheads for massive MIMO FDD systems. Finally, simulations are carried out to verify the performance of the proposed schemes and compared to its conventional counterparts with fewer pilots overhead, the proposed schemes improve the performance of the system.

The paper titled "Generalized Complex Quadrature Spatial Modulation" proposes two generalizations of CQSM, namely, generalized CQSM with unique combinations (GCQSM-UC) and with permuted combinations (GCQSM-PC). These two generalizations perform close to CQSM or outperform it, depending on the system parameters. Also, the proposed schemes require much less transmit antennas to achieve the same spectral efficiency of CQSM, for instance, assuming 16-QAM, GCQSM-PC, and GCQSM-UC require 10 and 15 transmit antennas, respectively, to achieve the same spectral of CQSM which is equipped with 32 antennas.

The paper titled "A Compact Ku-Band Active Electronically Steerable Antenna with Low-Cost 3D T/R Module" presents a novel compact Ku-band active electronically steerable antenna array design with a low-cost and integrated T/R 3D module employed for airborne synthetic aperture radar (SAR) systems. The entire system adopts 3D multilayer technology with vertical interconnection to construct the hermetically packaging RF modules. By assembling different multifunctional modules into a whole multilayer board, the 3D T/R technique greatly improves the system integration and reduces implementation cost and size.

The paper titled "SUBBASE: An Authentication Scheme for Wireless Sensor Networks Based on User Biometrics" proposes a new method of security for a wireless sensor

network (WSN). The proposed technique, Secure User Biometric Based Authentication Scheme (SUBBASe), is based on the user biometrics for WSNs. It achieves a higher security level, as well as improved network performance.

The paper titled "Enabling Noninvasive Physical Assault Monitoring in Smart School with Commercial Wi-Fi Devices" presents Wi-Dog, a noninvasive physical assault monitoring scheme that enables privacy-preserving monitoring in ubiquitous circumstances. The key intuition is that Wi-Fi signals are easily distorted by human motions, and motion-induced signals could convey informative characteristics, such as intensity, regularity, and continuity. Specifically, to explicitly reveal the substantive properties of physical assault, the authors propose a set of signal processing methods for informative components extraction by selecting sensitive antenna pairs and subcarriers. Then a novel signal-complexity-based segmentation method is developed as a location-independent indicator to monitor targeted movement transitions. The Wi-Dog is implemented on commercial Wi-Fi devices and evaluated in real indoor environments.

The paper titled "A Double Adaptive Approach to Tackle Malicious Users in Cognitive Radio Networks" proposes a double adaptive thresholding technique in order to differentiate legitimate users from doubtful and malicious users. Prior to the double adaptive approach, the maximal ratio combining (MRC) scheme is utilized to assign weight to each user such that the legitimate users experience higher weights than the malicious users. Double adaptive threshold is applied to give a fair chance to the doubtful users to ensure their credibility. A doubtful user that fails the double adaptive threshold test is declared as a malicious user. The results of the legitimate users are combined at the fusion center by utilizing Dempster-Shafer (DS) evidence theory. Effectiveness of the proposed scheme is proved through simulations by comparing with the existing schemes.

The paper titled "Smart Real-Time Video Surveillance Platform for Drowsiness Detection Based on Eyelid Closure" proposes drowsiness detection in real-time surveillance videos by determining if a person's eyes are open or closed. As a first step, the face of the subject is detected in the image. In the detected face, the eyes are localized and filtered with an extended Sobel operator to detect the curvature of the eyelids. Once the curves are detected, concavity is used to tell whether the eyelids are closed or open. Consequently, a concave upward curve means the eyelid is closed whereas a concave downwards curve means the eye is open. The proposed method is also implemented on hardware in order to be used in real-time scenarios, such as driver drowsiness detection.

The paper titled "Underwater Wireless Sensor Networks: A Review of Recent Issues and Challenges" conducts a survey of UWSN regarding underwater communication channel, environmental factors, localization, media access control, routing protocols, and effect of packet size on communication. The presently available methodologies are compared and their pros and cons are discussed to highlight new directions of research for further improvement in underwater sensor networks.

The paper titled "An Angle Estimation Method for Monostatic MIMO Radar Based on RCC-FLOM Algorithm" solves the angle estimation problem of coherent sources in the impulse noise background; a conjugate rotation invariant subspace algorithm based on reduced order fractional lower order covariance matrix is proposed. According to the conjugate rotation invariant subspace, the coherent source is decohered. The Monte-Carlo experiments show that the proposed algorithm has the advantages of high estimation probability and low root mean square error in the case of low signal-to-noise ratio, compared with the existing FLOM-MUSIC algorithm and FLOM-Unitary ESPRIT algorithm.

The paper titled "A Novel Study to Predict Trends and Policies for Mobile Communication in Multienvironment Regions" explores the usage of mobile communication to analyze the trends; for this, the authors select Khyber Pakhtunkhwa (KPK), Pakistan, a multienvironment based on farmers, students, employees, and businessmen. A field survey is carried out by designing a detailed structural questionnaire for viable collection of data to have superior analysis of different multienvironment based classes. A statistical analysis is performed through hypotheses and chi-square test on a large dataset based on sufficient number of observations collected through primary survey for each class. The survey results are provided in number of graphical and numerical illustrations to predict the mobile usage trends, evidently and effectively, of the people of KPK.

The paper titled "Flexible Queuing Model for Number of Active Users in Cognitive Radio Network Environment" presents a Soft Queuing Model (SQM) for number of active users present in a cognitive radio network (CRN) at some given instant. It starts with the existing cellular network where the upper limit for the number of channels and active users is well defined. The idea is then extended to the complicated scenario of CRNs where the upper limit is not deterministic for both the number of channels and the active users. Accordingly a probabilistic SQM is proposed under the condition that the number of channels and active cognitive users are both random variables. The proposed model will be useful to offer the level of reliability to the clients connected with CRN and hence to offer secure communication even on a cooperative CRN. The proposed model has been verified theoretically and simulations have been carried out in diversified set-ups to evaluate the performance.

The paper titled "Architecture for Collision-Free Communication Using Relaxation Technique" proposes an idea for collision problem in which network relaxation technique is used which is based on fast clique detection. The proposed approach results in high throughput in terms of latency and complexity. Furthermore, the proposed solution is able to solve the collision problem by connecting network optimization for achieving high throughput.

The paper titled "One-to-Many Relationship Based Kullback Leibler Divergence against Malicious Users in Cooperative Spectrum Sensing" proposes a technique based on Kullback Leibler Divergence (KLD) algorithm for mitigating the MUs attack in CSS. The secondary users (SUs) inform FC about the primary user (PU) spectrum availability by sending received energy statistics. Unlike the previous KLD

algorithm where the individual SU sensing information is utilized for measuring the KLD, in this work MUs are identified and separated based on the individual SU decision and the average sensing statistics received from all other users. The proposed KLD assigns lower weights to the sensing information of MUs, while the normal SUs information receives higher weights. The proposed method has been tested in the presence of always yes, always no, opposite, and random opposite MUs.

The paper titled "Backtracking Search Optimization Paradigm for Pattern Correction of Faulty Antenna Array in Wireless Mobile Communications" presents viable, simple, and accurate stochastic solver based on Backtracking Search Optimization Algorithm (BSA) for the pattern correction of faulty antenna array in mobile communication systems. A fitness function is developed to optimize the weights of the remaining healthy antenna elements in the array. The fitness function consists of two parts; the first part is based on mean square error approach for the reduction of sidelobes level, while, in the second part, steering vectors are used for the repositioning of nulls. Simulation results establish the validity of the BSA from its counterparts based on genetic algorithm and its memetic combination with pattern search technique.

Conflicts of Interest

There is no conflict of interest

Fawad Zaman
Sungchang Lee
Mohamad K. A. Rahim
Sarmadullah Khan

References

[1] F. Zaman, "Joint angle-amplitude estimation for multiple signals with L-structured arrays using bioinspired computing," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9428196, 12 pages, 2017.

[2] Y. Xiao, J. Xie, L. Huang, and H. C. So, "Multiantenna assisted source detection in toeplitz noise covariance," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 813–817, 2019.

[3] S. Akbar, F. Zaman, M. Asif, A. U. Rehman, and M. A. Raja, "Novel application of FO-DPSO for 2-D parameter estimation of electromagnetic plane waves," *Neural Computing and Applications*, pp. 1–10, 2018.

[4] K. Luo and A. Manikas, "Joint transmitter–receiver optimization in multitarget mimo radar," *IEEE Transactions on Signal Processing*, vol. 65, no. 23, pp. 6292–6302, 2017.

[5] H. Guo, Z. Yang, Y. Zou et al., "Secure beamforming for cooperative wireless-powered networks with partial CSI," *IEEE Internet of Things Journal*, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 9046409, 17 pages
https://doi.org/10.1155/2019/9046409

Research Article

Backtracking Search Optimization Paradigm for Pattern Correction of Faulty Antenna Array in Wireless Mobile Communications

Fawad Zaman,[1] **Hammad ul Hassan**,[1] **Shafqat Ullah Khan**,[2] **Ata ur Rehman**,[1] **Muhammad Asif Zahoor Raja**,[1] **and Shahab Ahmad Niazi**[3]

[1]*Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Attock, Pakistan*
[2]*Department of Electronics, The University of Buner, Buner, Khyber Pakhtunkhwa, Pakistan*
[3]*Department of Electronic Engineering, UCE&T, The Islamia University of Bahawalpur, Pakistan*

Correspondence should be addressed to Fawad Zaman; fawad@ciit-attock.edu.pk

Received 19 September 2018; Revised 7 February 2019; Accepted 17 February 2019; Published 11 July 2019

Academic Editor: Zhou Su

The demand for wireless mobile communication is growing exponentially with expectations that in the near future mobile device or user will be available in every corner of the globe. Alternatively, it increases the importance of antenna arrays which are responsible for transmission and reception of information. Every antenna array is projected to generate a desired pattern and, hence, failure of any antenna causes misrepresentation of the overall pattern in terms of increased side lobe levels and displacement of nulls from their original position. The aim of the study is to present viable, simple, and accurate stochastic solver based on backtracking search optimization algorithm (BSA) for the pattern correction of faulty antenna array in mobile communication systems. A fitness function is developed to optimize the weights of the remaining healthy antenna elements in the array. The fitness function consists of two parts: the first part is based on mean square error approach for the reduction of sidelobes level, while, in the second part, steering vectors are used for the repositioning of nulls. Simulation results establish the validity of the BSA from its counterparts based on genetic algorithm and its memetic combination with pattern search technique.

1. Introduction

The interest of researchers in wireless mobile communication systems is growing day by day due to several reasons. First the price for production of antenna arrays is getting cheaper and secondly, their size is getting smaller that ultimately decreases the power consumption due to which battery capacity is increasing. In order to improve the directivity (gain), usually an array is used that consists of two or more sensors or antennas. Antenna array is one of the main components that improve the performance and spectral efficiency of mobile communication systems [1–3]. Every antenna has its own weight that can be configured to get desired excitation level of that antenna. These weights have two parameters, namely, phase and amplitude, that can be adjusted individually. The combined effects of all weights generate the desired pattern having suitable side lobe level, nulls position, and null depths [1]. This process is known as array synthesis which is used for controlling amplitudes, positions, and phases to achieve the desired radiation patterns and eliminate noise and unwanted signals. Nulls can be used to achieve this goal by moving them in the direction of unwanted signals but this is not an easy task as direction of unwanted signal is not known or the signal's direction is changing on regular basis. This is known as nulling operation that can be achieved by adjusting amplitude-only [4], position-only [5], and phase-only [6] of the individual antenna to generate nulls in the direction of unwanted signals.

The antennas used in the array of mobile communication systems are expected to work for longer periods of time. Due to this longer operation of times, wear and tear or failure probability is also increased. For this reason, the antenna array must be able to evaluate its performance and detect possible breakdowns as the desired pattern is destroyed due

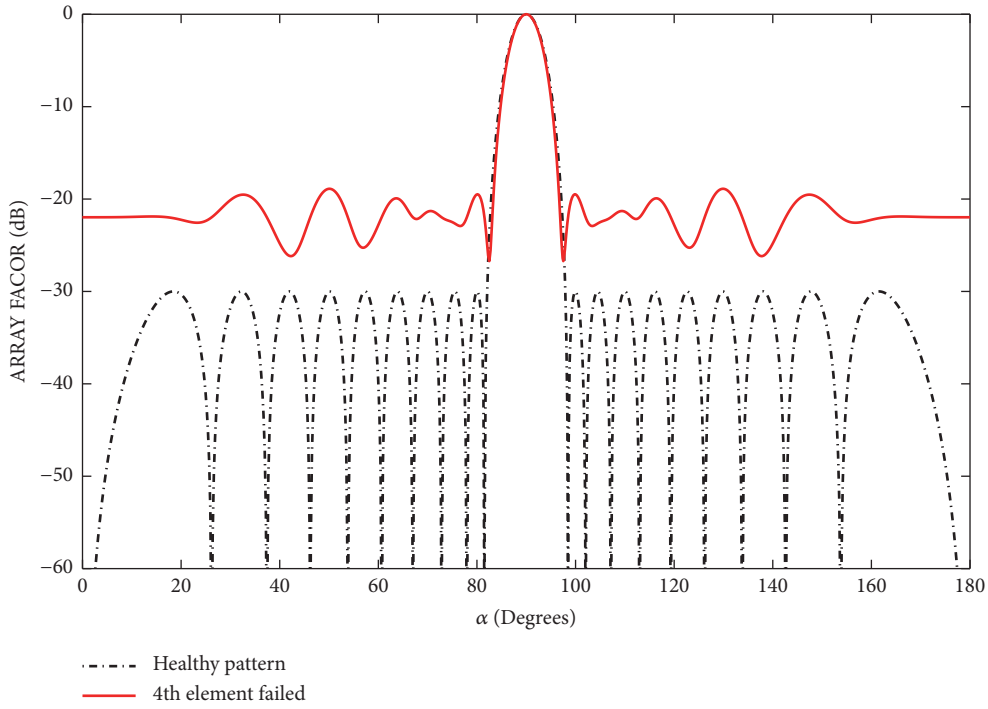


Figure 1: Array factor of 20-element antenna array vs. 4th element failed.

to the failure of single or multiple antennas of an array, which leads to widening the main beam, high side lobe levels, misplacement of nulls from their original position, and diminished null depth. [7, 8]. The effects of antenna failure on the array pattern are shown in Figure 1. An array must be able to compensate for the failure and maintain an error-free operation, as repairing a failed antenna is costly, as well as time consuming.

1.1. Related Work. In this modern and fast growing era of science and technology, meta-heuristic based optimization algorithms have their own importance and effectiveness that got applications in each field of science and engineering. A deterministic model yields a single solution of an experiment that is given some inputs but at the same time meta-heuristic based model gives a distribution of possible outputs with a probability of how likely each is to happen. They are conceptually simple, more flexible, and prominent particularly in the existence of local minima. The various metaheuristic techniques are particle swarm optimization [9], bee's algorithm [10], bacterial foraging algorithm [11], differential evolution [12], genetic algorithm (GA) [13], ant colony optimization [14], backtracking search optimization algorithm (BSA) [15], artificial bee colony algorithm [16], touring ant colony algorithm [17], firefly algorithms [18], etc. These techniques are implemented for many applications such as astrophysics, plasma and atomic physics, nonlinear optics, random matrix theory, electromagnetic, nanotechnology, fluid dynamics, electric machines, bioinformatics, energy, power, signal processing, controls, and communication; see [19–28] and the references cited therein. Among them recently BSA has attracted the attention of researcher due to its simplicity and efficiency. BSA has a single control parameter unlike other metaheuristic algorithms. Moreover, the problem solving capability of BSA is less dependent on the initial value of the parameter. The simple structure of the BSA enables it to solve multimodal functions effectively and quickly. In reported study [15], BSA has been statistically compared with PSO, ABC, self-adaptive differential evolution algorithm (SADE), adaptive differential evolution algorithm (JDE), covariance matrix adaptation evolution strategy (CMAES), and comprehensive learning particle swarm optimizer (CLPSO), and comparison was done using 3 real-world and 75 constrained benchmarks. In general, the simulations and comparisons in [15] show that BSA yielded more success as compared to the other algorithms. These facts motivate authors to explore in BSA for pattern correction of faulty antenna array in wireless mobile communications.

1.2. Innovative Contributions. Failure of any antenna element in the mobile communication systems causes distortion of the overall pattern in terms of increased side lobe levels and displacement of nulls from their original position of the transmitting beam. The exploration and exploitation of metaheuristic is investigated in the present study for pattern correction of faulty antenna array with the following salient features in terms of innovative contributions.

(i) The novel application of evolutionary computational heuristics through backtracking search optimization algorithm is presented for the correction of faulty array for wireless mobile communication systems.

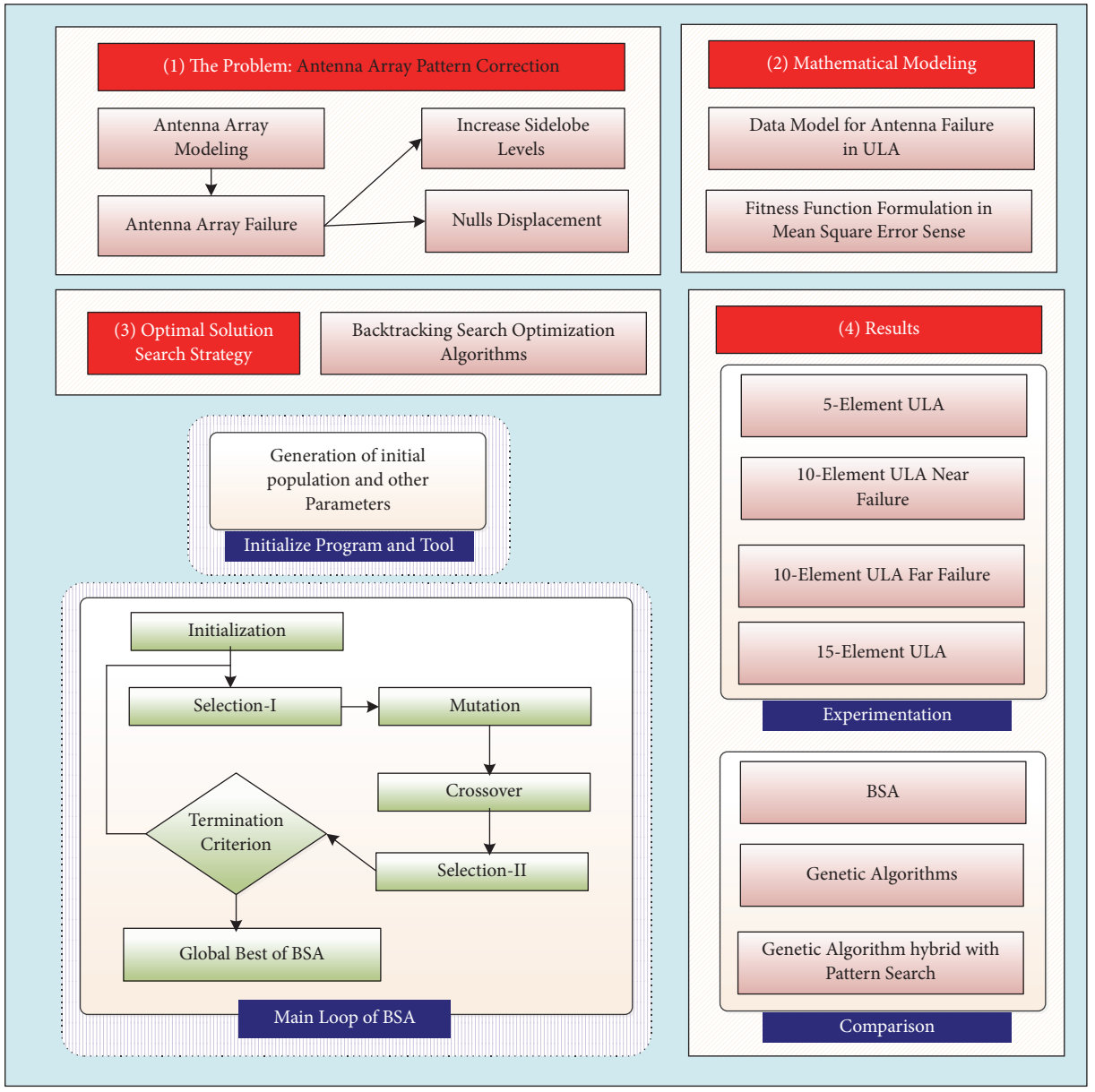


Figure 2: Schematic diagram of the proposed scheme.

(ii) In case of failure of any antenna element, a fitness function is constructed to optimize the weights of the remaining healthy antenna elements in the array. The developed fitness function consists of two parts, i.e., the first part is based on mean square error approach dedicated for the reduction of sidelobe levels while the second part is built on steering vectors used for the repositioning of nulls.

(iii) The simulation results show the validity of the proposed computational paradigm of BSA in terms of decreased sidelobe levels and nulling pattern. The comparison of BSA results from its counterparts based on GA and memetic combination of GA with pattern search (PS) further indorsed its superior performance.

1.3. Organization. The remainder of the paper is divided into four sections. In Section 2.1, the problem formulation is given for the pattern correction of uniform linear array (ULA) used for mobile communication systems. A detailed explanation for BSA is presented in Section 2.2 while Section 3 is dedicated for simulation results. Finally, the conclusions and future work directions are given in Section 4.

2. Design Methodology

The design scheme presented here consists of two parts: in the first part a date model for ULA is provided briefly while in the second part an overview of the optimization procedure based on backtracking search optimization methodology is provided. The schematic of the proposed methodology is shown in Figure 2.

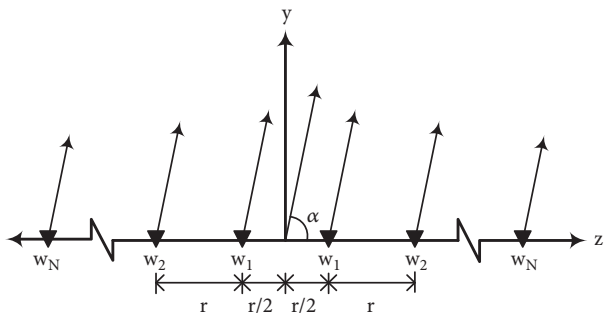


Figure 3: Uniform linear array of even number of sensors.

2.1. Problem Formulation. Let us consider ULA used for mobile communication which is composed of 2N (even) isotropic antenna placed symmetrically along z-axis, as shown in Figure 3. r represents inter-element spacing and N is the number of antennas, α is the elevation angle, while w_i is the complex weight corresponds to i^{th} antenna in the ULA.

The array factor for this ULA can be written as

$$\begin{aligned}(AF)_{2N} = {} & w_N e^{+j((2N-1)/2)kr\cos\alpha} + \cdots + w_2 e^{+j(3/2)kr\cos\alpha} \\ & + w_1 e^{+j(1/2)kr\cos\alpha} + w_1 e^{-j(1/2)kr\cos\alpha} \\ & + w_2 e^{-j(3/2)kr\cos\alpha} + \cdots \\ & + w_N e^{-j((2N-1)/2)kr\cos\alpha}\end{aligned} \tag{1}$$

where k represents the wavenumber of the signal. Simplifying (1) by applying Euler's identity gives

$$(AF)_{2N} = \sum_{i=1}^{N} w_i \cos\left[\frac{(2i-1)}{2} kr\cos\alpha\right] \tag{2}$$

By substituting $\eta = (\pi r\cos\alpha)/\lambda$, we get

$$(AF)_{2N} = \sum_{i=1}^{N} w_i \cos\left[(2i-1)\eta\right] \tag{3}$$

Similarly for odd number of elements, array factor can be calculated as

$$(AF)_{2N+1} = \sum_{i=1}^{N+1} w_i \cos\left[2(i-1)\eta\right] \tag{4}$$

In order to steer nulls and main beam in the desired direction, we need to calculate the weights adaptively.

Weighting Method: research community has proposed different techniques to choose the best suited weight for each antenna element to place the null(s) and/or main beam in a desired direction. For many practical applications the weights of ULA are calculated using Dolph-Tschebyscheff method [25, 26] and so we did in this work. Dolph-Tschebyscheff array was developed by Dolph using Tschebycheff polynomial so it is referred as Dolph-Tschebyscheff array. From (3) and (4), it can be observed that the array factor is the summation of N or 2N cosine terms, largest one being less than total number of antenna elements. These cosine terms can be written as a series of cosine functions as

$$\begin{aligned}
&n = 0 \longrightarrow \cos(n\vartheta) = 1 \\
&n = 1 \longrightarrow \cos(n\vartheta) = \cos\vartheta \\
&n = 2 \longrightarrow \cos(n\vartheta) = \cos(2\vartheta) = 2\cos^2\vartheta - 1 \\
&n = 3 \longrightarrow \cos(n\vartheta) = \cos(3\vartheta) = 4\cos^3\vartheta - 3\cos\vartheta \\
&n = 4 \longrightarrow \cos(n\vartheta) = \cos(4\vartheta) = 8\cos^4\vartheta - 8\cos^2\vartheta + 1 \\
&n = 5 \longrightarrow \cos(n\vartheta) = \cos(5\vartheta) \\
&\quad = 16\cos^5\vartheta - 20\cos^3\vartheta + 5\cos\vartheta
\end{aligned} \tag{5}$$

Using Euler's formula $[e^{j\vartheta}]^n = (\cos\vartheta + j\sin\vartheta)^n = e^{jn\vartheta} = \cos(n\vartheta) + j\sin(n\vartheta)$ and the trigonometric identity $\sin^2\vartheta = 1 - \cos^2\vartheta$ and letting $x = \cos\vartheta$, we get

$$\begin{aligned}
&n = 0 \longrightarrow \cos(n\vartheta) = 1 = T_0(x) \\
&n = 1 \longrightarrow \cos(n\vartheta) = x = T_1(x) \\
&n = 2 \longrightarrow \cos(n\vartheta) = 2x^2 - 1 = T_2(x) \\
&n = 3 \longrightarrow \cos(n\vartheta) = 4x^3 - 3x = T_3(x) \\
&n = 4 \longrightarrow \cos(n\vartheta) = 8x^4 - 8x^2 + 1 = T_4(x) \\
&n = 5 \longrightarrow \cos(n\vartheta) = 16x^5 - 20x^3 + 5x = T_5(x)
\end{aligned} \tag{6}$$

Equation (6) shows that each cosine term is related to a different polynomial $T_q(x)$; it should be noted that these relationships are only valid in the range $-1 \le x \le +1$ because $|\cos(n\vartheta)| \le 1$. The values for the Tschebyscheff polynomial can be found by the recursion formula:

$$T_q(x) = 2xT_{q-1}(x) - T_{q-2}(x) \tag{7}$$

Equation (7) shows that the next order polynomial can be calculated if the previous two polynomials are known. Numerous research articles have been written for effective calculation of Dolph-Tschebyscheff weights [29, 30].

The critical parameters for a ULA radiation pattern include null positioning, null depth, main beam amplitude, beam width, and side lobe level. In case, one or more elements get faulty due to any reason, like thunder storm and aging, the pattern of ULA is largely affected. If repair is not immediately available there is another possibility to reassign weights to the remaining healthy antennas. In order to recover the side lobe level and null positioning, a metaheuristic algorithm known as BSA is proposed in this work. The formulation of cost function is explained in the subsequent section.

Cost Function: in order to achieve pattern correction, a cost function is formulated. The aim is to obtain an array pattern, from a failed array that is close to that of healthy array by reconfiguring the weights of the remaining healthy elements. The cost function is given by

$$c = c_1 + c_2 \tag{8}$$

```
(1)   Initialization
Repeat
     (2) Selection – I
      Generation of trial population
          (3) Mutation
          (4) Crossover
      end
      (5) Selection – II
Until stopping criteria are met
```

Algorithm 1: Algorithm of BSA.

Table 1: BSA parameter settings.

Parameter	Definition	Value
Popsize	Population size	80
Dim	Dimension	9
DIM_RATE	Dimension rate	1
Epoch	Maximum iteration number	1000

Here c_1 denotes the function which is used to minimize side lobe levels (SLL). It is given by

$$c_1 = \sum \left(|SLL|_d\right)^2 = \sum_{i=1}^{p} \left|AF_H\left(\alpha_i\right) - AF_d\left(\alpha_i\right)\right|^2. \quad (9)$$

Equation (9) represents the mean square error (MSE) in order to minimize the SLL. The reduction is achieved by using MSE between $AF_H(\alpha_i)$, which is healthy level and $AF_d(\alpha_i)$ which is faulty level. Here, p is the limit up to which the side lobe level is to be minimized.

In (8) c_2 denotes the function which is used to generate nulls in the desired locations. Initial theory about null placement was presented by Steyskal et al. [31] in which a steering vector based null positioning is proposed. Using this idea we need the following equation for null generation in the required direction:

$$AF\left(\alpha_i\right) = \mathbf{w}^H \mathbf{v}\left(\alpha_i\right) = 0 \quad (10)$$

where $\mathbf{w}$ is a $N \times 1$ vector, defined as

$$\mathbf{w} = \left[w_{-M}, \cdots, w_{-1}, w_0, w_1, \cdots, w_M\right]^T \quad (11)$$

and $\boldsymbol{v}(\alpha_i)$ is $N \times 1$ vector, used as steering vector

$$\mathbf{v}\left(\alpha_i\right) = \left[e^{-j((N-1)/2)kr\cos\alpha_i}, e^{-j((N-3)/2)kr\cos\alpha_i} \cdots e^{j((N-1)/2)kr\cos\alpha_i}\right]^T_{N\times 1} \quad (12)$$

N is the number of elements. The constraint matrix $\mathbf{A}$ is defined by following equation:

$$\mathbf{A} = \left[\mathbf{v}\left(\alpha_1\right), \mathbf{v}\left(\alpha_2\right), \ldots, \mathbf{v}\left(\alpha_{M0}\right)\right] \quad (13)$$

Here α_i for $i=1,2,\ldots M_o$ represents the direction of the nulls. To get the nulls in a desired direction the below-mentioned squared weighed function should be minimized:

$$c_2 = \left\|\mathbf{w}^H \mathbf{A}\right\|^2 \quad (14)$$

Backtracking search optimization (BSA) is used to minimize c as given in (8), which is the overall cost function.

2.2. Backtracking Search Optimization Algorithm. BSA is a newly developed evolutionary algorithm (EA) presented in 2013 by Pinar Civicioglu [15]. The BSA has an ability to efficiently handle various types of optimization problems. It uses the information from the previously generated population to generate a new population having superior fitness values. BSA is divided into five main steps:

(1) Initialization
(2) Selection-I
(3) Mutation
(4) Crossover
(5) Selection-II

The basic or general working of BSA is shown in Algorithm 1, while the details of the fundamental steps in term of pseudocode are given in Pseudocode 1.

3. Simulation Results

In this section, performance of BSA is validated and several simulations are performed for a single antenna failure in 5, 10, and 15 elements ULA, respectively. The effects of antenna failure near and far from the center are also considered. Throughout the simulations, the inter-element spacing 'r' is taken to be 0.5λ. The performance of BSA is compared to already available techniques for array pattern correction using GA and GA-PS as reported in [1]. The parameter setting for BSA is given in Table 1.

One can observe that failure of element near the center (4^{th} element) causes more damage to the pattern in terms of high side lobe level and displacement of nulls from their original position as compared to failure of element far from center (9^{th}). These effects are shown in Figures 4 and 5 which

Step 1: Initialization:

An initial population P is initialized as

$$\mathbf{P}_{kl} \sim V(lower_l, upper_l) \tag{15}$$

Where $k = 1, 2, \cdots, N$, and $l = 1, 2, \cdots, D$, V is the uniform distribution function, N is population size, and D is problem dimension. P_k represents k^{th} chromosome in the population P. $lower_l$ and $upper_l$ represents the lower limit upper limit of the solution respectively.

Step 2: Selection-I:

In Selection-I, historical population P^{his} is determined, which is used to calculate the direction of search for optimum solution. Initial P^{his} is determined by:

$$\mathbf{P}^{his}{}_{kl} \sim \mathrm{V}(lower_l, upper_l)\ . \tag{16}$$

This process is done through the "if-then" rule as:

$$if\ c < d\ then\ \mathbf{P}^{his}{}_{kl} \coloneqq \mathbf{P} \mid c, d \sim \mathrm{V}(0, 1) \tag{17}$$

Here $\coloneqq$ is defined as update operation, c and d are random numbers in the range [0,1] which help decide if the Phis is selected from the former generation. After this step the order of the chromosomes is shuffled by random shuffling function given by:

$$\mathbf{P}^{his} \coloneqq \mathrm{permuting}(\mathbf{P}^{his}) \tag{18}$$

Step 3: Mutation:

A mutation is done in $\mathbf{P}^{his}$ to generate a trail population $\mathbf{P}^{mutant}$ using:

$$\mathbf{P}^{mutant} = \mathbf{P} + G(\mathbf{P}^{his} - \mathbf{P}) \tag{19}$$

So, $\mathbf{P}^{mutant}$ is due to the propagation of chromosomes of $\mathbf{P}$ in the direction set by $(\mathbf{P}^{his} - \mathbf{P})$ and G controls amplitude of $(\mathbf{P}^{his} - \mathbf{P})$. Use of $\mathbf{P}^{his}$ gives a partial advantage to BSA of the experiences of previous population in creating an intelligent trial population $\mathbf{P}^{mutant}$.

Step 4: Crossover:

It creates the final trial population $\mathbf{P}^T$, the initial step towards $\mathbf{P}^T$ was $\mathbf{P}^{mutant}$ set by (19). Target population is improved by chromosomes of $\mathbf{P}^T$ having improved fitness. For this crossover uses two procedures. First determines a binary matrix map $\mathbf{M}$ of $N \times D$ order. Its function is to indicate particles of $\mathbf{P}^T$ that are to be changed by using relevant particles of $\mathbf{P}$. The initial value of $\mathbf{M}_{m,n}$ is set as 1, and trial population $\mathbf{P}^T$ is updated as:

$$\mathbf{P}^T{}_{n,m} \coloneqq \mathbf{P}_{n,m} \tag{20}$$

Here $n \in \{1, 2, \cdots, N\}$ & $m \in \{1, 2, \cdots, D\}$. Crossover strategy of BSA is presented in Algorithm 2. BSA uses a unique crossover strategy as compared to other EAs. The mix-rate parameter controls the number of particles that are mutated in a trial by using $[mixrate \times rnd \times D]$, is mentioned in line (3) of Algorithm 2. This function is the main reason that BSA crossover is unique to other EAs

BSA's $\mathbf{M}$ is defined by two predefined strategies being used randomly. The first strategy implements mix-rate (Algorithm 2 line (2) - (4)), and other allows one chromosome to mutate, chosen at random in each trial (Algorithm 2 line (6) - (8)).

The mutation strategy results in overflow of some particles of the PT during crossover procedure. For this, a boundary control process is defined that keeps the individual inside the bounds. Its algorithm is given in Algorithm 3.

Step 5: Selection-II:

In this process the particles of $\mathbf{P}^T$ having improved fitness are updated by their corresponding better particles of $\mathbf{P}$. The global minimum and global minimizer are also updated based on the best fitness so far. The particle having best fitness is called global minimizer denoted by Pbest and its fitness is called global minimum.

Step 6: Termination

Finally, optimization process of the BSA is stopped if one the following criterion is met:

(i) Max number of iteration is reached, or

(ii) The fitness value is below a certain threshold

Pseudocode 1: Pseudocode of BSA with procedural details.

```
     Input: P^mutant , mix-rate, N and D
     Output: P^T : trial population
(0)  M_(1:N,1:D) = 1
(1) If c < d | c, d ~ V(0,1) then
(2)         for k from 1 to N do
(3)               M_k,v(1:[mixrate × rnd × D]) = 0 | v = permuting(<1,2,...,D>)
(4)         end
(5)  else
(6)         for k from 1 to N do
(7)               M_k,randi(D) = 0
(8)         end
(9)  end
(10) P^T := P^mutant
(11) for k from 1 to N do
(12)        for l from 1 to D do
(13)              if M_k,l = 1 then
(14)                  P^T_k,l := P_k,l
(15)        end
(16) end
```

Algorithm 2: Crossover strategy of BSA.

```
     Input : P^T, Search space limits (i.e., lower_l , upper_l )
     Output : P^T
(0) for k from 1 to N do
(1)      for l from 1 to D do
(2)           If ( P^T_k,l < lower_l ) or ( P^T_k,l > upper_l ) then
(3)                P^T_k,l = rnd × ( upper_l − lower_l ) + lower_l
(4)           End
(5)      End
(6) End
```

Algorithm 3: Boundary mechanism of BSA.

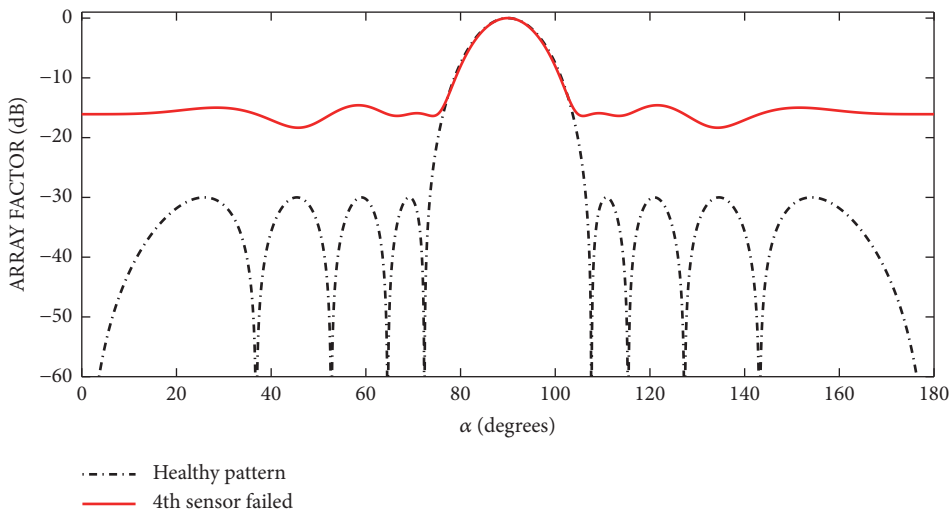


Figure 4: Pattern of 10-element array with 4^{th} element failure.

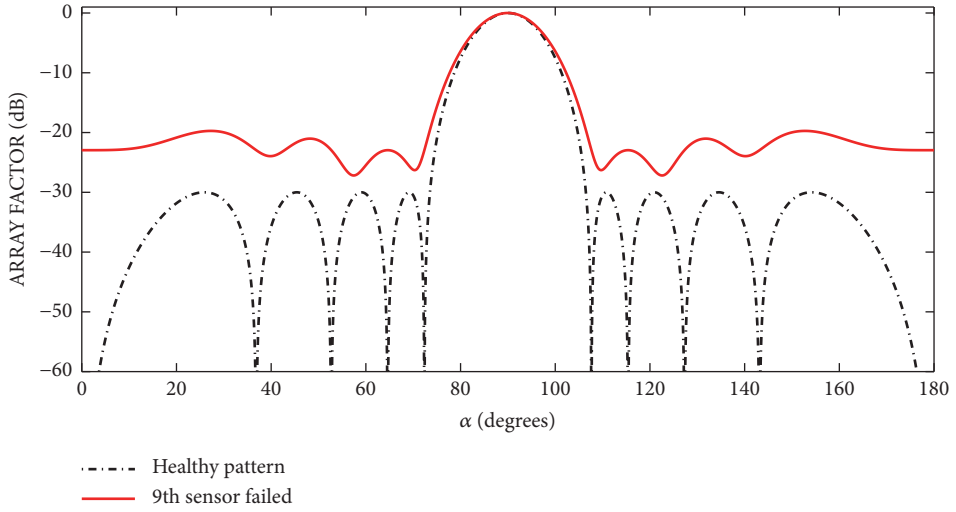


Figure 5: Pattern of 10-element array with 9^{th} element failure.

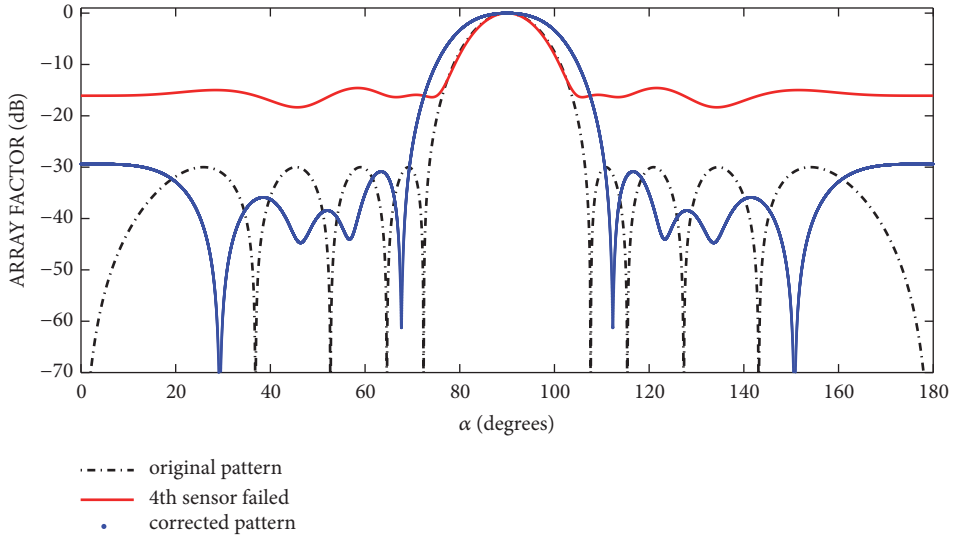


Figure 6: SLL recovered for 10-element 4^{th} sensor failed array.

also verify that BSA is an efficient scheme for solving both cases.

The examples of first group are simulated using a 10-element array with element failure near center. The simulations show the recovery of SLL in Figure 6, recovery of single null in Figures 7 and 8, and recovery of multiple nulls in Figures 9–11.

The second group of examples is simulated using a 10-element array with element failure far from center. The simulations show the recovery of SLL in Figure 12, recovery of single null in Figures 13 and 14, and recovery of multiple nulls in Figures 15 and 16.

The third group of examples is simulated using a 5-element ULA. The simulations show the recovery of SLL in Figure 17 and recovery of single nulls in Figure 18. The fourth group of examples is simulated using a 15-element array with element failure far from center. The simulations show the recovery of SLL in Figure 19 and recovery of multiple nulls in Figures 20–22. Tables 2, 3, 4, and 5 show the comparison between faulty pattern and recovered pattern for each case, respectively.

Performance of BSA is compared in terms of efficiency and accuracy, with an already proven technique in [1]. The results presented in Table 6 show the comparison results. The table shows that BSA is much quicker than both GA and GA-PS due to its structural simplicity, BSA is also more accurate. The simulations for GA and GA-PS were also done on the same computer.

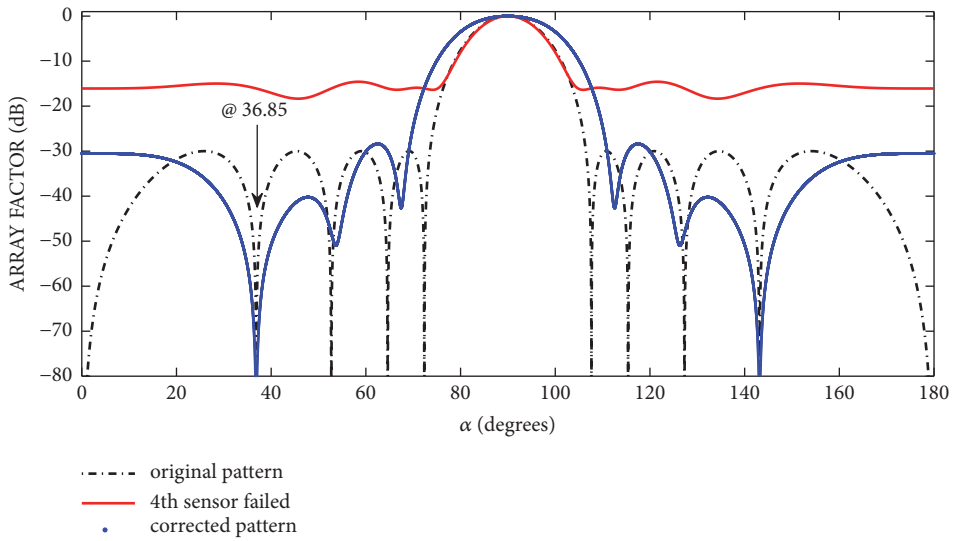


Figure 7: Fourth null recovered for 10-element 4^{th} sensor failed array.

Table 2: Comparison of faulty and corrected pattern for 10-element array with failure near center.

Faulty pattern		Corrected pattern		Recovery of nulls
SLL(dB)	*NDL (dB)*	*SLL(dB)*	*NDL (dB)*	
-16	*-18*	-30	-85	1 null
-16	*-18*	-29	-105	2 nulls
-16	*-16*	-29	-98	
-16	*-18*	-28	-111	3 nulls
-16	*-17*	-28	-108	
-16	*-16*	-28	-122	
-16	*-18*	-30	-68	4 nulls
-16	*-17*	-30	-103	
-16	*-16*	-30	-72	
-16	*-16*	-30	-105	

Table 3: Comparison of faulty and corrected pattern for 10-element ULA with failure far from center.

Faulty pattern		Corrected pattern		Recovery of nulls
SLL(dB)	*NDL (dB)*	*SLL(dB)*	*NDL (dB)*	
-16	*-18*	-31	-108	1 null
-16	*-18*	-28	-100	3 nulls
-16	*-17*	-28	-105	
-16	*-16*	-28	-117	
-16	*-18*	-30	-80	4 nulls
-16	*-17*	-30	-93	
-16	*-16*	-30	-95	
-16	*-16*	-30	-105	

Table 4: : Comparison of faulty and corrected pattern for 5-element array.

Faulty pattern		Corrected pattern		Recovery of nulls
SLL(dB)	*NDL (dB)*	*SLL(dB)*	*NDL (dB)*	
-19	*-18*	-28	-78	2 nulls
-19	*-16*	-30	-81	

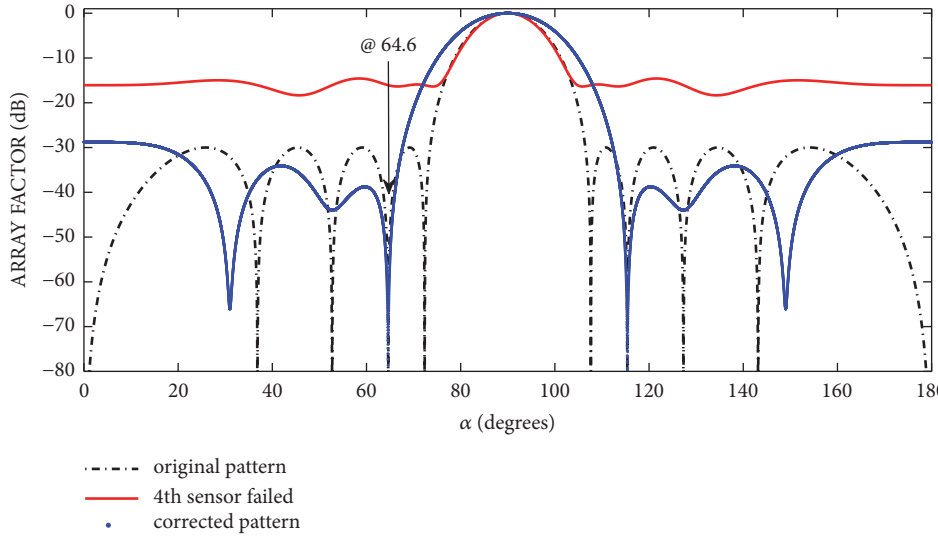


Figure 8: Second null recovered for 10-element 4^{th} sensor failed array.

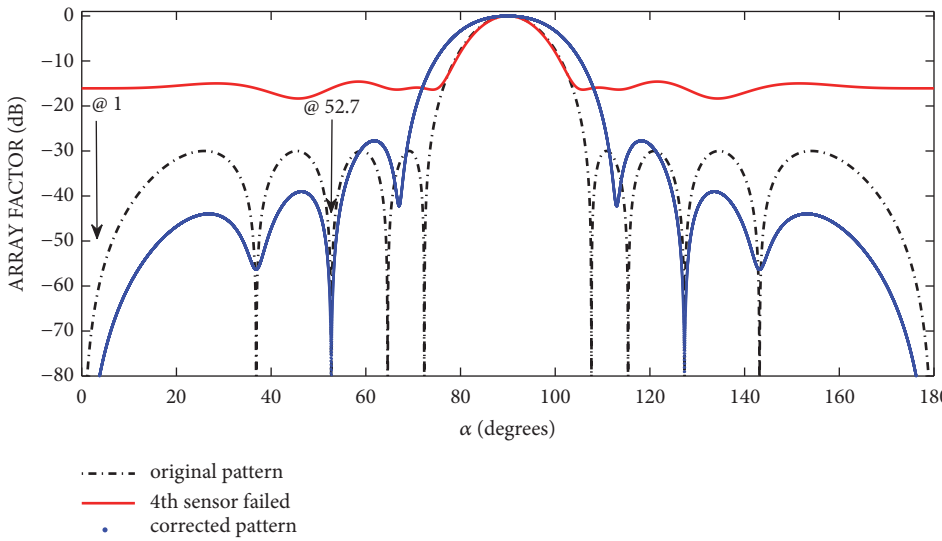


Figure 9: 3rd and 5^{th} null recovered for 10-element ULA 4^{th} sensor failed.

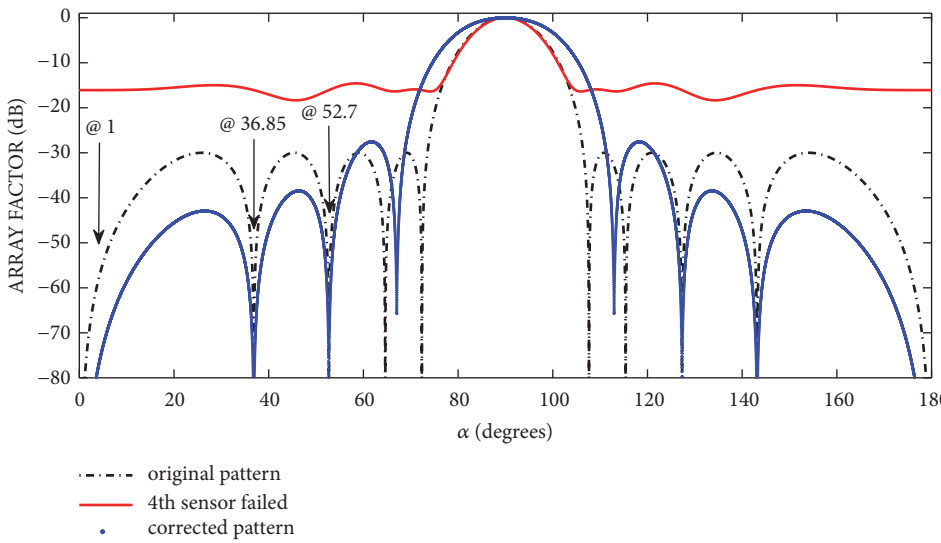


Figure 10: 3^{rd}, 4^{th}, and 5^{th} null recovered for 10-element ULA 4^{th} sensor failed.

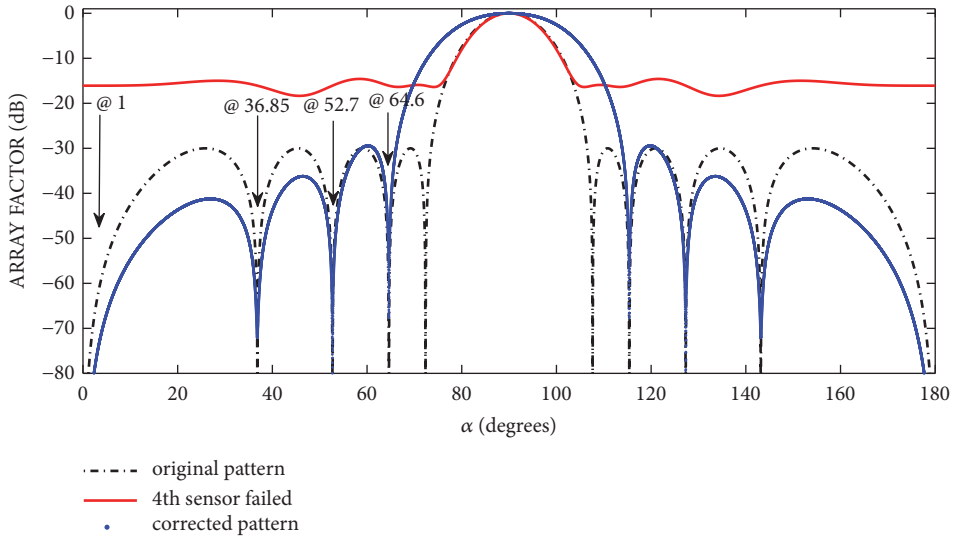


Figure 11: 2nd, 3rd, 4th, and 5th null recovered for 10-element ULA 4^{th} sensor failed.

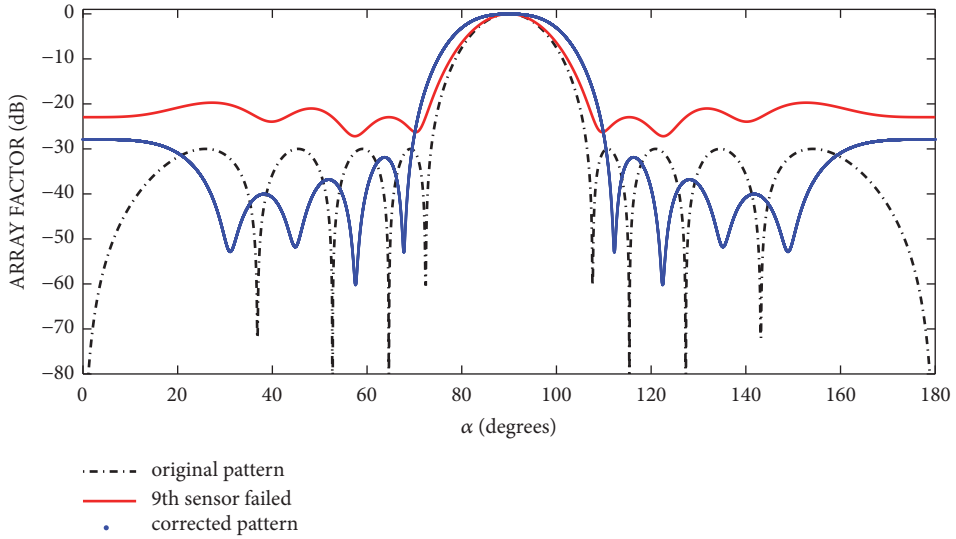


Figure 12: SLL recovered for 10-element 9^{th} sensor failed array.

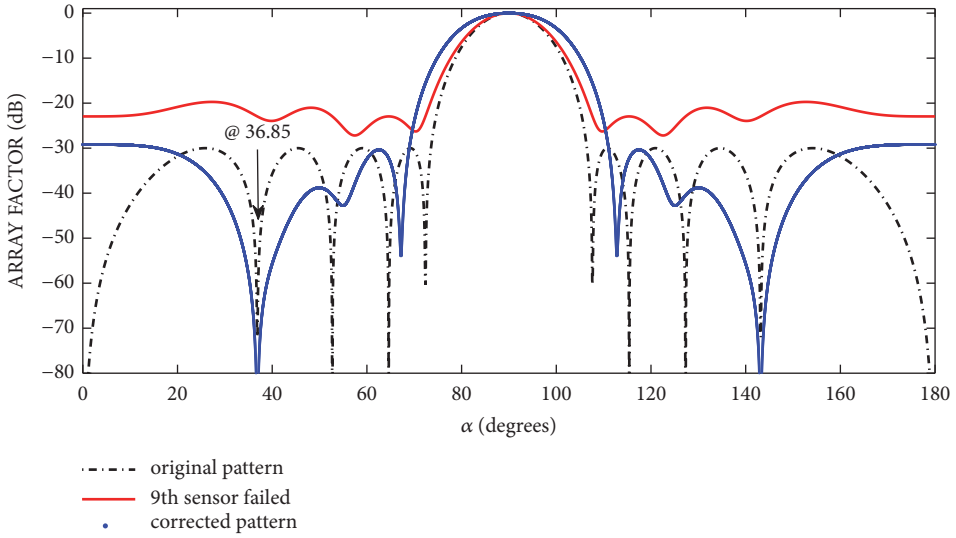


Figure 13: 4th null recovered for 10-element ULA 9^{th} sensor failed.

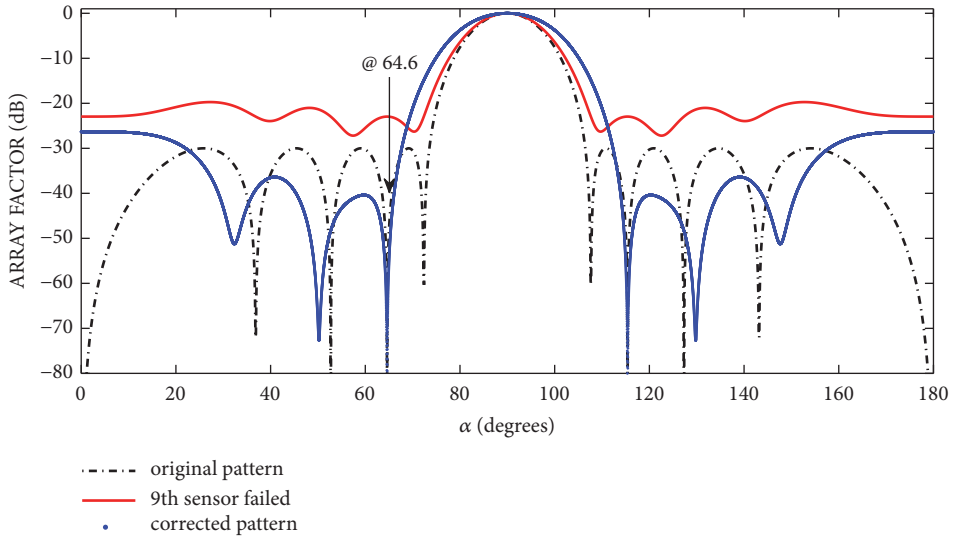


Figure 14: 2nd null recovered for 10-element ULA 9^{th} sensor failed.

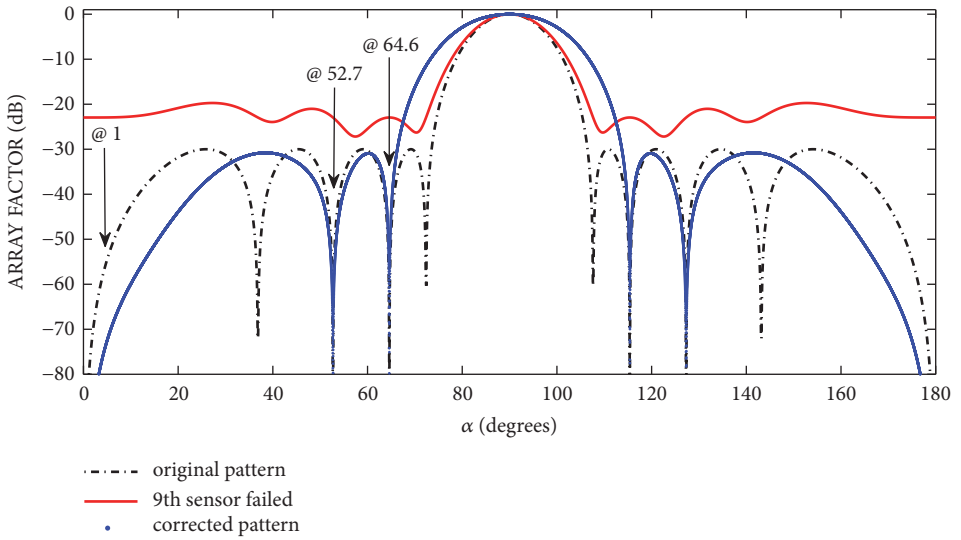


Figure 15: 2nd, 3rd, and 5th null recovered for 10-element ULA 9^{th} sensor failed.

Table 5: Comparison of faulty and corrected pattern for 15-element array.

Faulty pattern		Corrected pattern		Recovery of nulls
SLL(dB)	*NDL (dB)*	*SLL(dB)*	*NDL (dB)*	
-10	*-16*	-30	-96	2 nulls
-10	*-16*	-30	-111	
-10	*-19*	-29	-61	4 nulls
-10	*-14*	-29	-91	
-10	*-12*	-29	-95	
-10	*-12*	-29	-108	
-10	*-18*	-30	−41	7 nulls
-10	*-15*	-30	-55	
-10	*-14*	-30	-71	
-10	*-19*	-30	-64	
-10	*-12*	-30	-82	
-10	*-10*	-30	-60	
-10	*-14*	-30	-71	

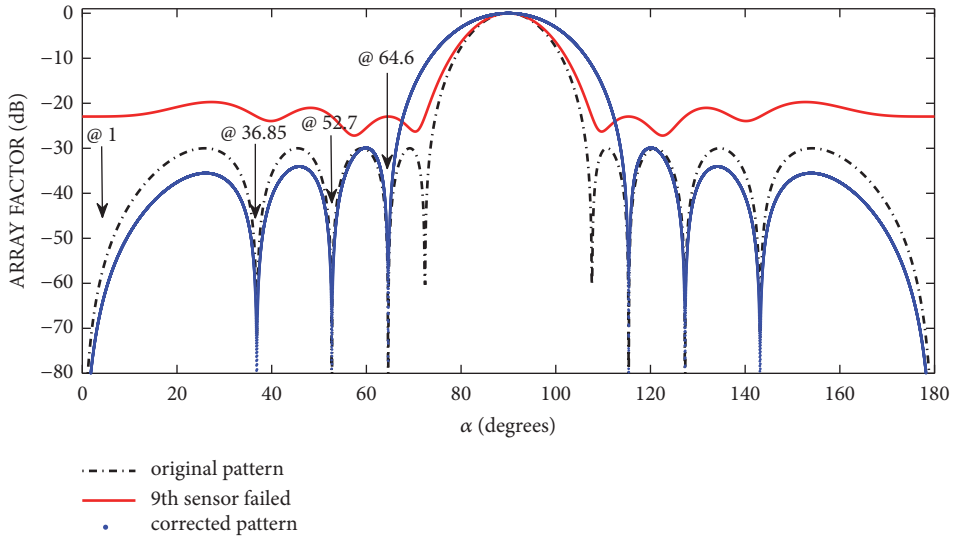


FIGURE 16: 2nd, 3rd, 4th, and 5th null recovered for 10-element ULA 9^{th} sensor failed.

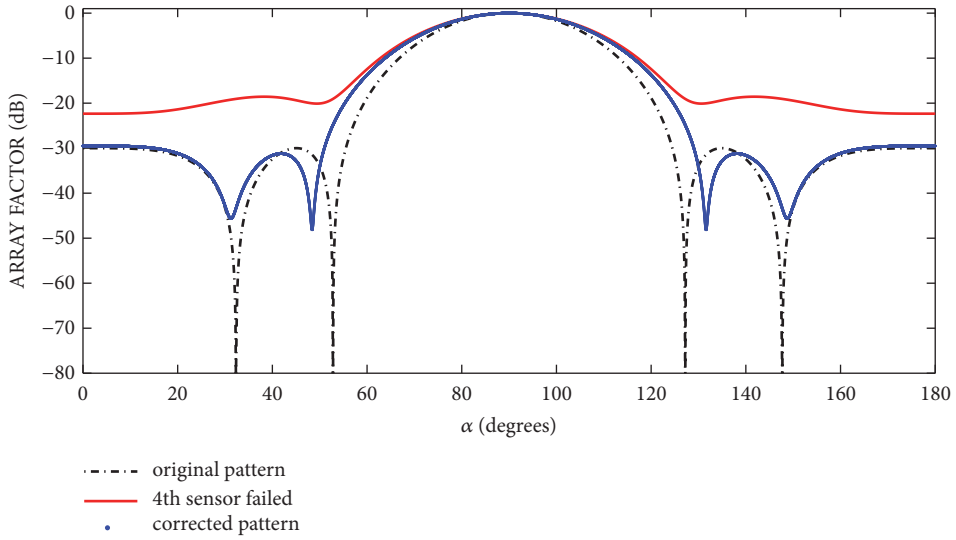


FIGURE 17: SLL recovered for 5-element 4^{th} sensor ULA failed.

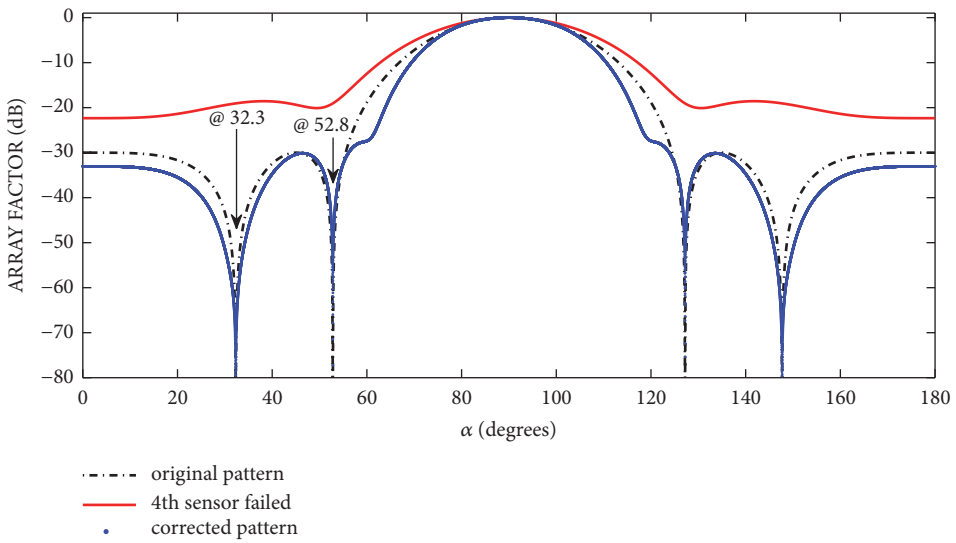


FIGURE 18: 1^{st} and 3^{rd} recovered for 5-element 4^{th} sensor failed array.

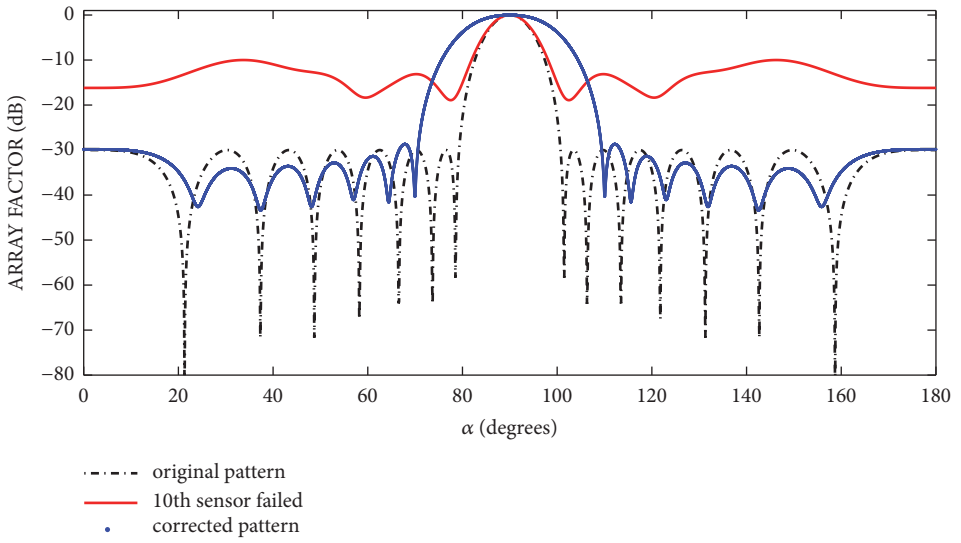


Figure 19: SLL recovered for 15-element ULA 10^{th} sensor failed.

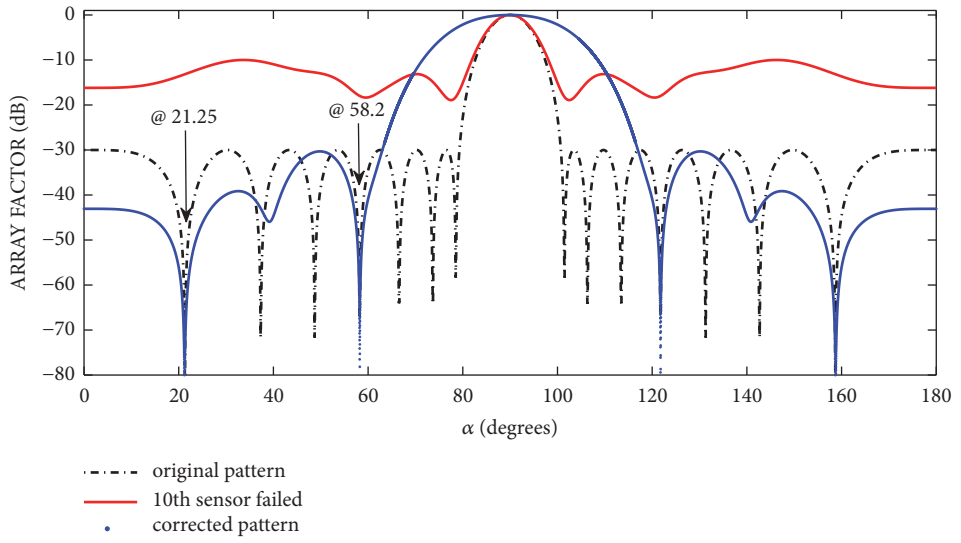


Figure 20: 4^{th} and 7^{th} null recovered for 15-element ULA 10^{th} sensor failed.

Table 6: Comparison of BSA, GA, and GA-PS results.

BSA		GA [1]		GA-PS [1]		No of array elements
Fitness value (10^{-3})	*Time elapse* (s)	*Fitness value* (10^{-3})	*Time elapse* (s)	*Fitness value* (10^{-3})	*Time elapse* (s)	
1.57	49.79	6.24	147.1	5.41	154.3	5-element
4.56	74.15	8.56	204.5	6.43	218.9	10-element near failure
3.27	73.8	7.68	202.7	4.29	217.2	10-element far failure
7.08	94.3	9.86	295.4	8.47	335.8	15-element

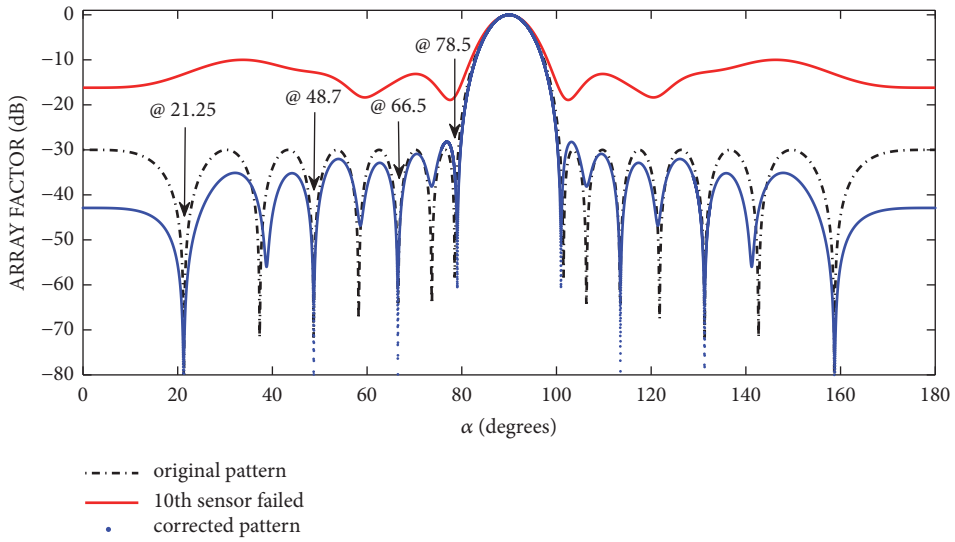


FIGURE 21: 1st, 3rd, 5th, and 7th null recovered for 15-element ULA 10^{th} sensor failed.

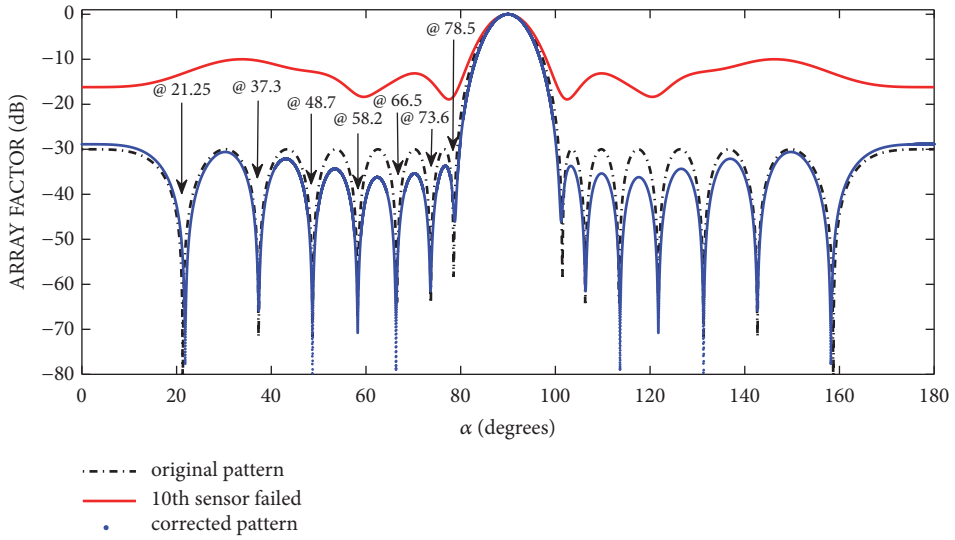


FIGURE 22: 1st, 2nd, 3rd, 4th, 5th, 6th, and 7th null recovered for 15-element ULA 10^{th} sensor failed.

4. Conclusion and Future Work

In this paper, strength of BSA is used for the correction of faulty antenna array in wireless mobile communication systems. BSA has successfully reduced side lobe levels and repositioned single and multiple nulls at the original positions when any antenna element in the ULA is failed. The comparison of BSA is made with GA and GA-PS and it has been established that BSA produced better results in terms of computational complexity.

In future, one can explore BSA for other geometries of antenna arrays such as L shape, planar, and circular arrays. Similarly, BSA can be successfully applied to other problems in array signal processing such as adaptive beamforming, parameter estimation of signals, and tracking.

Data Availability

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

Ethical Approval

All the authors of the manuscript declared that there is no research involving human participants and/or animal.

Disclosure

The authors are responsible for any fees, if any, associated with this manuscript.

Conflicts of Interest

All the authors of the manuscript declared that there are no potential conflicts of interest.

Authors' Contributions

Fawad Zaman contributed in conceptualization, supervision, and review and editing, Hammad ul Hassan contributed in formal analysis and writing original draft, Shafqat Ullah Khan contributed in investigation and writing review and editing, Ata ur Rehman, contributed in formal analysis and validation, Muhammad Asif Zahoor Raja contributed in methodology and software, and Shahab Ahmad Niazi contributed in visualization and review and editing.

References

[1] S. U. Khan, I. M. Qureshi, F. Zaman, and A. Naveed, "Null placement and sidelobe suppression in failed array using symmetrical element failure technique and hybrid heuristic computation," *Progress in Electromagnetics Research*, no. 52, pp. 165–184, 2013.

[2] T. E. Tuncer and B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation*, Academic Press, 2009.

[3] S. Akbar, M. A. Z. Raja, F. Zaman, T. Mehmood, and M. A. R. Khan, "Design of bio-inspired heuristic techniques hybridized with sequential quadratic programming for joint parameters estimation of electromagnetic plane waves," *Wireless Personal Communications*, vol. 96, no. 1, pp. 1475–1494, 2017.

[4] T. Pavani, G. S N Raju, and P. V Sridevi, "Pattern synthesis of linear array with position and position-amplitude control using firefly algorithm," *International Journal of Computer Applications*, vol. 99, no. 7, pp. 18–23, 2014.

[5] S. K. Mahto and A. Choubey, "A novel hybrid IWO/WDO algorithm for interference minimization of uniformly excited linear sparse array by position-only control," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 250–254, 2016.

[6] A. Kumar, R. Mudumbai, S. Dasgupta et al., "A scalable feedback mechanism for distributed nullforming with phase-only adaptation," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 1, pp. 58–70, 2015.

[7] W. P. M. N. Keizer, "Amplitude-only low sidelobe synthesis for large thinned circular array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 1157–1161, 2012.

[8] Z.-U. Khan, A. N. Malik, F. Zaman, S. A. Hussain, and A.-R. Khan, "Sidelobe suppression with null steering by independent weight control," *International Journal of Antennas and Propagation*, vol. 2015, pp. 1–6, 2015.

[9] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*, pp. 760–766, Springer, USA, 2011.

[10] B. Yuce, E. Mastrocinque, M. S. Packianather, A. Lambiase, and D. T. Pham, "The bees algorithm and its applications," *Handbook of Research on Artificial Intelligence Techniques and Algorithms*, pp. 122–151, 2014.

[11] K. M. Passino, "Bacterial foraging optimization," in *Innovations and Developments of Swarm Intelligence Applications*, vol. 219, 2012.

[12] V. Feoktistov, *Differential evolution*, Springer, USA, 2006.

[13] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, no. 2, pp. 65–85, 1994.

[14] M. Dorigo and M. Birattari, "Ant colony optimization," in *Encyclopedia of machine learning*, pp. 36–39, Springer, Boston, MA, USA, 2011.

[15] P. Civicioglu, "Backtracking search optimization algorithm for numerical optimization problems," *Applied Mathematics and Computation*, vol. 219, no. 15, pp. 8121–8144, 2013.

[16] W. F. Gao and S. Y. Liu, "A modified artificial bee colony algorithm," *Computers & Operations Research*, vol. 39, no. 3, pp. 687–697, 2012.

[17] D. Karaboga, K. Guney, and A. Akdagli, "Antenna array pattern nulling by controlling both amplitude and phase using modified touring ant colony optimization algorithm," *International Journal of Electronics*, vol. 91, no. 4, pp. 241–251, 2004.

[18] X.-S. Yang, "Firefly algorithms for multimodal optimization," in *International Symposium on Stochastic Algorithms*, pp. 169–178, Springer, Berlin, Heidelberg, Germany, 2009.

[19] M. A. Z. Raja, M. A. Manzar, F. H. Shah, and F. H. Shah, "Intelligent computing for Mathieu's systems for parameter excitation, vertically driven pendulum and dusty plasma models," *Applied Soft Computing*, vol. 62, pp. 359–372, 2018.

[20] Z. Sabir, M. A. Manzar, M. A. Raja, M. Sheraz, and A. M. Wazwaz, "Neuro-heuristics for nonlinear singular Thomas-Fermi systems," *Applied Soft Computing*, vol. 65, pp. 152–169, 2018.

[21] S. Akbar, F. Zaman, M. Asif, A. U. Rehman, and M. A. Z. Raja, "Novel application of FO-DPSO for 2-D parameter estimation of electromagnetic plane waves," *Neural Computing and Applications*, pp. 1–10, 2018.

[22] A. Mehmood, A. Zameer, and M. A. Z. Raja, "Intelligent computing to analyze the dynamics of Magnetohydrodynamic flow over stretchable rotating disk model," *Applied Soft Computing*, vol. 67, pp. 8–28, 2018.

[23] A. Zameer, J. Arshad, A. Khan, and M. A. Z. Raja, "Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks," *Energy Conversion and Management*, vol. 134, pp. 361–372, 2017.

[24] F. Zaman, "Joint angle-amplitude estimation for multiple signals with L-structured arrays using bioinspired computing," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9428196, 12 pages, 2017.

[25] M. Pakdaman, A. Ahmadian, S. Effati, S. Salahshour, and D. Baleanu, "Solving differential equations of fractional order using an optimization technique based on training artificial neural network," *Applied Mathematics and Computation*, vol. 293, pp. 81–95, 2017.

[26] A. Jafarian, R. Jafari, A. K. Golmankhaneh, and D. Baleanu, "Solving fully fuzzy polynomials using feed-back neural networks," *International Journal of Computer Mathematics*, vol. 92, no. 4, pp. 742–755, 2015.

[27] M. A. Zahoor Raja, Z. Shah, M. Anwaar Manzar, I. Ahmad, M. Awais, and D. Baleanu, "A new stochastic computing paradigm for nonlinear Painlevé II systems in applications of random matrix theory," *The European Physical Journal Plus*, vol. 133, no. 7, p. 254, 2018.

[28] A. Jafarian, M. Mokhtarpour, and D. Baleanu, "Artificial neural network approach for a class of fractional ordinary differential equation," *Neural Computing and Applications*, vol. 28, no. 4, pp. 765–773, 2017.

[29] A. D. Bresler, "A new algorithm for calculating the current distributions of Dolph- Tschebyscheff arrays," *IEEE Transactions on Antennas and Propagation*, vol. 28, no. 6, pp. 951-952, 1980.

[30] N. Balakrishnan and R. Sethuraman, "Easy generation of Dolph-Tschebyscheff excitation coefficients," *Proceedings of the IEEE*, pp. 1508-1509, 1981.

[31] H. Steyskal, R. A. Shore, and R. L. Haupt, "Methods for null control and their effects on the radiation pattern," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 404–409, 1986.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 6374764, 15 pages
https://doi.org/10.1155/2019/6374764

Research Article

Compressive Sensing Based Channel Estimation for Massive MIMO Communication Systems

Athar Waseem,[1] **Aqdas Naveed**,[1] **Sardar Ali**,[2] **Muhammad Arshad**,[3] **Haris Anis**,[1] **and Ijaz Mansoor Qureshi**[4]

[1]*International Islamic University, Islamabad, Pakistan*
[2]*University of Buner, Pakistan*
[3]*University of Peshawar, Pakistan*
[4]*AIR University, Islamabad, Pakistan*

Correspondence should be addressed to Athar Waseem; athar.waseem@iiu.edu.pk

Received 9 November 2018; Revised 13 March 2019; Accepted 21 April 2019; Published 27 May 2019

Guest Editor: Sungchang Lee

Massive multiple-input multiple-output (MIMO) is believed to be a key technology to get 1000x data rates in wireless communication systems. Massive MIMO occupies a large number of antennas at the base station (BS) to serve multiple users at the same time. It has appeared as a promising technique to realize high-throughput green wireless communications. Massive MIMO exploits the higher degree of spatial freedom, to extensively improve the capacity and energy efficiency of the system. Thus, massive MIMO systems have been broadly accepted as an important enabling technology for 5th Generation (5G) systems. In massive MIMO systems, a precise acquisition of the channel state information (CSI) is needed for beamforming, signal detection, resource allocation, etc. Yet, having large antennas at the BS, users have to estimate channels linked with hundreds of transmit antennas. Consequently, pilot overhead gets prohibitively high. Hence, realizing the correct channel estimation with the reasonable pilot overhead has become a challenging issue, particularly for frequency division duplex (FDD) in massive MIMO systems. In this paper, by taking advantage of spatial and temporal common sparsity of massive MIMO channels in delay domain, nonorthogonal pilot design and channel estimation schemes are proposed under the frame work of structured compressive sensing (SCS) theory that considerably reduces the pilot overheads for massive MIMO FDD systems. The proposed pilot design is fundamentally different from conventional orthogonal pilot designs based on Nyquist sampling theorem. Finally, simulations have been performed to verify the performance of the proposed schemes. Compared to its conventional counterparts with fewer pilots overhead, the proposed schemes improve the performance of the system.

1. Introduction

Multiple-input multiple-output (MIMO) systems have multiple antennas at both the transmitter and receiver ends. By addition of multiple antennas, higher degree of freedom in wireless channels (in terms of time and frequency dimensions) can be obtained in order to achieve target of high data rates. For this reason, major performance progress can be attained in terms of system reliability and both spectral and energy efficiency. And also such higher degrees of freedom can be exploited using beamforming given that channel state information is available. There are large number of antenna elements (around tens or even hundreds) deployed at both sides, the transmitter and receiver. It is important to note that the transmit antennas may be distributed or colocated in different applications. Also, the huge number of receive antennas can be acquired by one device or distributed to many devices [1, 2]. Additionally, massive MIMO systems help in minimizing the effects of noise and fast fading, and also intracell interference can be reduced using straightforward detection and linear precoding methods. By appropriately implementing multiuser MIMO (MU-MIMO) in massive MIMO systems, the design of medium access control (MAC) layer can be more simplified by getting rid of complicated scheduling algorithms [2]. With MU-MIMO, signals to individual users can be easily separated using the

same time-frequency resource, at base station. Therefore, these main advantages make it feasible to introduce the massive MIMO system as a promising candidate for 5G wireless communication networks.

One of the major issues in massive MIMO systems is the accurate acquisition of the channel state information (CSI) for beamforming, resource allocation, signal detection, etc. Due to large antennas placed at the BS, the estimation of channels linked with hundreds of transmit antennas is required at users which results in high pilot overhead. Hence, the precise channel estimation with the low pilot overhead is a challenging task [2, 3].

There are several research challenges which are required to be solved before massive MIMO can be fully integrated into future wireless systems [4, 5]:

(i) Beamforming will involve a huge amount of channel state information, which will be challenging for the downlink.

(ii) MIMO can be unfeasible for FDD systems, but can be employed in time division duplex (TDD) systems due to having channel reciprocity.

(iii) Conventional channel estimation approaches require a large pilot and feedback overhead, which typically scales proportionally with number of BS transmit antennas, which results in unfeasible condition for large-scale FDD MIMO systems.

(iv) In addition, massive MIMO experiences pilot contamination from other cells if the transmit power is high; otherwise it will be affected by thermal noise. Methods will be required to solve this in order to deliver the promised performance.

(v) There is a need for channel models of massive MIMO systems, without which it will be difficult for researchers to perfectly validate the algorithms and techniques.

(vi) For channel estimation, TDD scenarios are only considered for massive MIMO due to the excessive cost of channel estimation and feedback. Even for TDD to work, massive MIMO channel calibration can prove to be a big achievement. New methods and schemes will be needed for the purpose of channel estimation in massive MIMO systems.

(vii) Extremely fast processing algorithms will be required for processing the massive amount of data from the RF chains.

The above-mentioned challenges have been addressed up to some extent in the last few years. To date, many researches on massive MIMO avoided the challenge of considering FDD systems by simply assuming the TDD protocol. The uplink CSI can be more easily obtained at the BS due to the less number of single-antenna users and the strong capability of processing at BS. And then by leveraging the channel reciprocity property, the CSI in the downlink can be directly tacked [6, 7]. However, due to the fact that radio frequency chains suffer from calibration error and restricted coherence time, the CSI obtained in the uplink may not be correct for the downlink [8, 9]. Additionally FDD systems have low latency as compared to TDD; therefore the communication is more efficient [10]. Thus, it is of significance to discover the challenging issue of channel estimation for FDD massive MIMO systems, which can assist massive MIMO to be backward attuned with existing FDD dominated cellular networks. There has been extensive analysis on channel estimation for traditional small-scale FDD MIMO systems [11, 12]. It has been established that the equally spaced and equally powered orthogonal pilots can be ideally suitable to estimate the noncorrelated Rayleigh MIMO channels for single OFDM symbol, where the pilot overhead required increases as the number of transmit antennas increases [11]. The pilot overhead to estimate Rician MIMO channels can be compacted by exploiting the spatial correlation of MIMO channels [13, 14]. Furthermore, by taking advantage of the temporal channel correlation, more reduced pilot overhead can be attained to estimate MIMO channels linked to multiple OFDM symbols [15–17]. Presently, orthogonal pilots have been extensively introduced in the current MIMO systems; due to small number of transmit antennas (i.e., up to eight antennas in LTE-Advanced system), they have reasonable pilot overhead [12, 18]. Though, this is a critical issue in massive MIMO systems due to massive figure of antennas at the BS (i.e., up to 128 antennas or even more at the BS [19]). An approach of exploiting the temporal correlation and the delay domain sparsity of channels for achieving the reduced pilot overhead has been presented for FDD massive MIMO systems in [20, 21], but with large number of transmit antennas, the interference cancellation of pilot sequences of different transmit antennas will be difficult. In [22–25] the spatial correlation in sparsity of delay domain MIMO channels is utilized to estimate channels achieving the reduced pilot overhead, but the assumption level of the known channel sparsity to the user is impractical. In [26–28], the compressive sensing (CS) based channel estimation schemes are presented by exploiting the spatial channel correlation; however due to having nonideal antenna array, the leveraged spatial correlation can be impaired [2, 8]. In [29] a novel doubly selective channel estimation scheme for MIMO-OFDM systems is proposed with both time and frequency-domain training (TFDT) based on structured compressive sensing (SCS). In [30] a new sparse channel model based on dictionary learning is proposed, which adapts to the cell characteristics and promotes a sparse representation. The learned dictionary is able to more robustly and efficiently represent the channel and improve downlink channel estimation accuracy. Furthermore, observing the identical AOA/AOD between the uplink and downlink transmission, a joint uplink and downlink dictionary learning and compressed channel estimation algorithm is proposed to perform downlink channel estimation utilizing information from the simpler uplink training, which further improves downlink channel estimation. In [31] a channel estimator based on block distributed compressive sensing (BDCS) is proposed for the large-scale MIMO systems. BDCS exploits structured sparsity to reduce the pilot overhead. In [32] a new way to estimate sparse channels by construction of beamforming

dictionary matrices is presented. Continuous basis pursuit (CBP) algorithm that exploits the sparse nature of channels to adaptively estimate the multipath mmWave channels is proposed. In [33] an open-loop and a closed-loop channel estimation scheme for massive MIMO are presented, but the channel statistics cannot be perfectly known to the user in the long term, whereas in conventional broadband wireless communication systems, delay domain channels basically exhibit the sparse nature due to the large channel delay spread and having limited number of major scatterers in the propagation environments [21, 34].

In the meantime, MIMO communication systems have similar scatterers in the transmission environment; therefore BS channels associated with one user and several transmit antennas experience similar type of path delays, which shows that these delay domain channels share common sparsity specially in the case of not having very large aperture of antenna array [2, 35]. Furthermore, throughout the coherence time, such sparsity is nearly unchanged which is due to the fact that path delays fluctuate at very slow rate as compared to path gains because of temporal correlation of channels [36]. In [37] CS based block sparsity adaptive matching pursuit (BSAMP) algorithm is proposed, which targets the estimation of channels of MIMO system with unknown number of channels paths. The BSAMP exploits the joint sparse nature of massive MIMO channels. In [38] CS based probability-weighted subspace pursuit (PWSP) algorithm is proposed, which exploits the probability information attained from previously estimated CIRs to recover the uplink channels in massive MIMO scenario. In [39] such properties of channels in MIMO are considered as the spatiotemporal common sparsity, which is generally ignored in current work. Finally, a general idea of the CS method is presented in [40], in which fundamental setup, recovery techniques, and guarantee of performance are discussed. Furthermore, different subproblems of CS, i.e., sparse approximation, identification of support, and sparse identification, are discussed, with respect to some applications of wireless systems. Design issues of wireless systems, limitations and potentials of CS algorithms, useful tips, and prior information are also discussed to achieve maximum performance. In [40] a valuable guidance is provided for researchers working in the area of wireless communication systems. Before getting into the detail of spatiotemporal common sparsity, we present an overview of compressive sensing.

In this paper we propose a CS based efficient nonorthogonal pilot scheme for massive MIMO communication systems by exploring the temporal and spatial sparsity of massive MIMO channels. And then we propose a channel estimation scheme, i.e., sparsity update CoSaMP (SUCoSaMP), which exploits the temporal and spatial sparsity of massive MIMO channels. The proposed pilot scheme is significantly different from the conventional schemes and substantially reduces the pilot overhead. The proposed pilot scheme employs fully identical subcarriers for pilots of several transmit antennas in a specific antenna group. The antennas placed at base station (BS) are subdivided into groups based on the observation that the coherence time of path gains is inversely proportional to the system carrier frequency whereas the variation in path delays is inversely proportional to the signal bandwidth. Therefore the decision of making antenna groups and determining the number of antennas to be included in one antenna group is taken according to the given system parameters, i.e., systems frequency, system bandwidth, and antenna spacing at BS. Furthermore, considering the antenna array geometry of BS, the proposed nonorthogonal pilot scheme is a space-time adaptive pilot scheme that adaptively changes its design according to the given system parameters. The proposed CS algorithm based channel estimation scheme SUCoSaMP considers the initial sparsity level as 1 and then regularly updates the sparsity level until the stopping criteria are met or a correct sparsity level is achieved in a scenario where sparsity level is unknown. Compared with the conventional CS algorithms SP and OMP and with other available CS based algorithms, the proposed CS based pilot and channel estimation scheme is tested through simulations on systems with different parameters. It was verified that the proposed schemes provided improved channel estimation performance.

The organization of the paper is as follows. Sections 2 and 3 present the details about delay domain spatial and temporal sparsity of massive MIMO communication channels, respectively. Section 4 presents the proposed nonorthogonal pilot scheme based on CS theory. Section 5 illustrates the CS based massive MIMO channel estimation scheme. Section 6 provides the simulation results. And in Section 7 we present the conclusion

Notation. Uppercase and lowercase boldface letters represent the matrices and vectors, respectively. The matrix transpose, inversion, and conjugate transpose are denoted by $(.)^T$, $(.)^{-1}$, and $(.)^*$, respectively. The Moore-Penrose inversion of matrix is denoted by $(.)^\dagger$. The l_2-norm operation and Frobenius-norm operation are denoted by $\|.\|_2$ and $\|.\|_F$, respectively. The cardinality of a set and the j-th column vector of the matrix $\boldsymbol{\psi}$ are denoted by $|.|_C$ and $\boldsymbol{\psi}^{(j)}$, respectively. Finally the complementary set of the set T and the integer floor operator are denoted by T^C and $\lfloor . \rfloor$, respectively.

2. Delay Domain Spatial Sparsity

Considerable experimental researches have discovered that, in delay domain massive MIMO, channels demonstrate spatial sparsity. This is due to the fact that the number of significant scatterers is limited in wireless communication in fading environments, while in communication between base station (BS) and users, the transmission distance is very large as compared to the distance between several antennas in an antenna array placed at BS. That is to say, CIRs associated with several transmit antennas and one user exhibit similar path delays; therefore they also share identical common support of CIRs.

Consider a massive MIMO-OFDM system where M transmit antennas are placed at BS. The CIR between m-th transmit antenna and one single-antenna user for z-th OFDM symbol is expressed by

$$\boldsymbol{d}_{z,m} = \left[d_{z,m}(1), d_{z,m}(2), \ldots, d_{m,r}(L)\right]^T \quad (1)$$

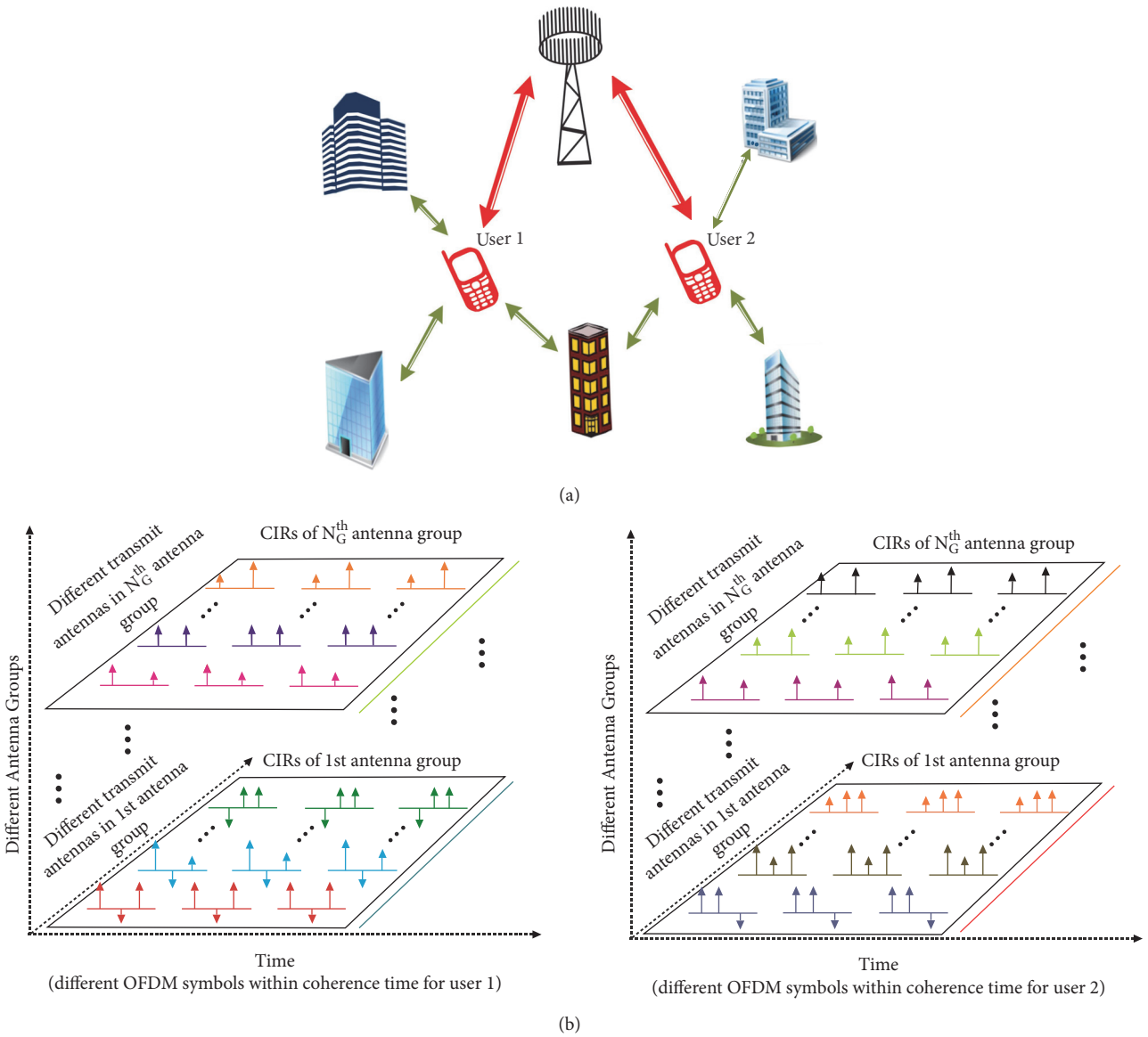


Figure 1: Delay domain spatial and temporal sparsity of massive MIMO channels: (a) limited number of scatterers and common scatterers in wireless communication scenario; (b) spatial and temporal sparsity of massive MIMO channels in delay domain between two users and colocated antenna array.

where $1 \leq m \leq M$ and L is the channel length equivalent to the maximum delay spread. Let $S_{z,m}$ be the sparsity level of CIR between one transmit-receive antenna pair, i.e., the number of nonzero elements in $\boldsymbol{d}_{z,m}$; the support of $\boldsymbol{d}_{z,m}$ can be expressed as

$$P_{z,m} = \text{supp}\{\boldsymbol{d}_{z,m}\} = \{l : |d_{z,m}[l]| > 0\} \quad \text{with } 1 \leq l \leq L \tag{2}$$

where $S_{z,m} = |P_{z,m}|_C$ fulfilling $S_{z,m} \ll L$. And due to spatial sparsity we have the following.

$$P_{z,1} = P_{z,2} = \cdots = P_{z,M} \tag{3}$$

The delay domain spatial sparsity with specific system parameters will be detailed in Section 5.

Figure 1 shows the common sparse pattern of CIRs for different transmit-receive antenna pairs.

3. Delay Domain Temporal Sparsity

In [39], the authors have discovered that fast time-varying channels illustrate temporal correlation. That is to say, the variations in path gains are significant whereas the path delays are almost invariant for various consecutive OFDM symbols. The reason is that path delays variation duration over time-varying channels is inversely proportional to the signal bandwidth while the coherence time of path gains is inversely proportional to carrier frequency of the system [28].

Consider a system with signal bandwidth $B = 10$ MHz and carrier frequency of system $f_c = 2$ GHz; the path delays vary much slower than the path gains [11]. Therefore, due to the nearly invariant path delays, the CIRs share the common sparsity for R adjacent OFDM symbols over the coherence time of path delays. The supports of CIRs associated with R successive OFDM symbols signify the following.

$$P_{1,m} = P_{z,m} = \cdots = P_{R,m}, \quad 1 \le m \le M \tag{4}$$

Figure 1 demonstrates the temporal sparsity of massive MIMO channels.

4. Proposed Nonorthogonal Pilot Scheme

In this section we shall propose a nonorthogonal pilot scheme based on the above-mentioned two observations. First the basic idea to divide the antennas placed at BS into subgroups is proposed. The proposed scheme of dividing the antennas into subgroups is based on system parameters and the temporal and spatial sparsity of massive MIMO channels. Then after creating the antenna groups, a specific nonorthogonal pilot scheme is proposed.

4.1. Proposed Scheme for Creating the Antenna Groups. There will be the uniform distribution of antennas to be included in subgroups. Consider a system with signal bandwidth B, system carrier frequency f_c, total number of antennas placed at the BS in a uniform linear antenna array M, number of subgroups N_g, and number of antennas in one group M_{ng}. The maximum resolvable distance D_{max} must be smaller than c/10B [38]; that is to say, to have two channel taps resolvable, the time interval of arrival must be less than 1/10B, where c is the speed of light [38]. The two successive antennas are spaced by the distance $d_m = \lambda/2$; therefore the maximum distance between two antennas can be $d_T = d_m(M - 1)$ and the maximum distance between two antennas in a single group can be $d_{M_{ng}} = d_m(M_{ng} - 1)$. In order to guarantee the spatial sparsity and based on the condition that $D_{max}/c < 1/10B$, the $d_{M_{ng}}$ must satisfy $d_{M_{ng}} < D_{max}$.

And the formula for number of antennas M_{ng} in each subgroup N_g can be derived as follows:

$$d_m\left(M_{ng} - 1\right) \le \frac{c}{10B} \tag{5}$$

$$\frac{\lambda}{2\left(M_{ng} - 1\right)} \le \frac{c}{10B} \tag{6}$$

$$M_{ng} \le \frac{c}{5B\left(c/f_c\right)} + 1 \tag{7}$$

which clearly shows that M_{ng} is a function of B and f_c. Therefore the N_g can be given by the following.

$$N_g = \left\lfloor \frac{M}{M_{ng}} \right\rfloor + \alpha \tag{8}$$

The constant α represents an integer to make sure the uniform distribution of antennas in each subgroup, i.e., M, is completely divisible by N_g. Similarly actual number of antennas M_{ngf} in each group can be calculated by the following.

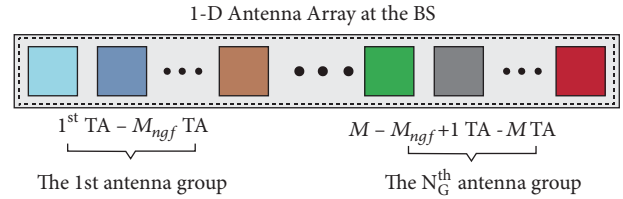


Figure 2: 1D antenna array at the BS.

$$M_{ngf} = \frac{M}{N_g} \tag{9}$$

The formula for N_g will ensure the spatial sparsity for any massive MIMO system with large 1D antenna array placed at BS. Moreover, since the M_{ng} is a function of system parameters B and f_c, if the B and f_c are changed, the N_g still ensures the spatial sparsity for antennas. And also N_g is the minimum number of subgroups a system must have to ensure spatial sparsity; moreover by increasing N_g, spatial sparsity of system will remain preserved. For example, consider a system with the following specifications:

$M = 128$ 1D antenna array, $f_c = 2$ GHz, B = 20 MHz, M_{ng} given in (7) can be calculated as 21, N_g will be equal to 8 with $\alpha = 2$, and M_{ngf} will be equal to 16. Based on these calculations the initial condition is satisfied, that is, $d_{M_{ng}} < D_{max}$; hence the spatial sparsity is preserved for the system.

The N_g antenna groups are shown in Figure 2, where TA denotes the transmit antenna.

In Section 4.2, first we will derive a massive MIMO system model. And then specific pilot design will be proposed in Section 4.3.

4.2. Massive MIMO System Model. Consider a massive MIMO-OFDM system with M transmit antennas placed at BS communicating with one user; ξ_n represents the index set of pilot subcarriers for n-th antenna group; the choice of ξ_n will be detailed in Section 4.3). There are total N subcarriers in one OFDM symbol, out of which N_p corresponds to pilot subcarriers and the pilot sequence of m-th transmit antenna in n-th antenna group is denoted by $\mathbf{s}_{\mathrm{m,n}} \in \mathbb{C}^{N_p \mathrm{x} 1}$. $\mathbf{y}_{\mathrm{z,n}} \in \mathbb{C}^{N_p \mathrm{x} 1}$ is the received vector of the pilot sequence of z-th OFDM symbol of n-th antenna group at user for N_g antenna groups; $\mathbf{y}_{\mathrm{z,n}}$ can be expressed as

$$\mathbf{y}_{\mathrm{z,n}} = \sum_{m=1}^{M_{ngf}} \mathrm{diag}\left\{\mathbf{s}_{\mathrm{m,n}}\right\} \mathbf{R}|_{\xi_n,\mathrm{n}} \begin{bmatrix} \mathbf{d}_{\mathrm{m,z,n}} \\ \mathbf{0}_{(N-L)\times 1} \end{bmatrix} + \mathbf{w}_{\mathrm{z,n}} \tag{10}$$

with $1 < \mathrm{n} < N_g$

$$\mathbf{y}_{\mathrm{z,n}} = \sum_{m=1}^{M_{ngf}} \mathbf{S}_{\mathrm{m,n}}\, \mathbf{R}_{\mathbf{L}}|_{\xi_n,\mathrm{n}}\, \mathbf{d}_{\mathrm{m,z,n}} + \mathbf{w}_{\mathrm{z,n}}$$

$$= \sum_{m=1}^{M_{ngf}} \boldsymbol{\psi}_{\mathrm{m,n}} \mathbf{d}_{\mathrm{m,z,n}} + \mathbf{w}_{\mathrm{z,n}} \tag{11}$$

where $\mathbf{R} \in \mathbb{C}^{N\mathrm{x}N}$ is the Discrete Fourier Transform matrix (DFT), $\mathbf{y}_{\mathrm{z,n}}$ is expressed after exclusion of guard interval, $\mathbf{d}_{\mathrm{m,z,n}} \in \mathbb{C}^{Lx1}$ is the CIR vector of $\mathrm{m_{th}}$ transmit antenna of n-th antenna group for $\mathrm{z_{th}}$ OFDM symbol, $\mathbf{R}|_{\xi_n,\mathrm{n}} \in \mathbb{C}^{N_p\mathrm{x}N}$ is a submatrix comprised of N_p rows selected according to ξ_n, $\mathbf{R_L}|_{\xi_n,\mathrm{n}} \in \mathbb{C}^{N_p\mathrm{x}L}$ contains the first L columns of $\mathbf{R}|_{\xi_n,\mathrm{n}} \in \mathbb{C}^{N_p\mathrm{x}N}$, $\mathbf{S}_{\mathrm{m,n}} = \mathrm{diag}\{\mathbf{s}_{\mathrm{m,n}}\} \in \mathbb{C}^{N_p\mathrm{x}N_p}$, $\boldsymbol{\psi}_{\mathrm{m,n}} = \mathbf{S}_{\mathrm{m,n}}\mathbf{R_L}|_{\xi_n,\mathrm{n}} \in \mathbb{C}^{N_p\mathrm{x}L}$, and $\mathbf{w}_{\mathrm{z,n}}$ represents the Additive White Gaussian Noise of n-th antenna group for z-th OFDM symbol.

The equation can be rearranged as

$$\mathbf{y}_{\mathrm{z,n}} = \boldsymbol{\psi}_{\mathrm{n}} \widetilde{\mathbf{d}}_{\mathrm{z,n}} + \mathbf{w}_{\mathrm{z,n}} \tag{12}$$

where $\boldsymbol{\psi}_{\mathrm{n}} = [\boldsymbol{\psi}_{1,\mathrm{n}}, \boldsymbol{\psi}_{2,\mathrm{n}}, \ldots, \boldsymbol{\psi}_{M_{ngf},\mathrm{n}}] \in \mathbb{C}^{Np\times M_{ngf}L}$ and the aggregate CIR vector of n-th antenna group is given by $\widetilde{\mathbf{d}}_{\mathrm{z,n}} = [\mathbf{d}_{1,\mathrm{z,n}}, \mathbf{d}_{2,\mathrm{z,n}}, \ldots, \mathbf{d}_{M_{ngf},\mathrm{z,n}}]^{\mathrm{T}} \in \mathbb{C}^{M_{ngf}L\times 1}$. As explained earlier, the CIR vector $\mathbf{d}_{\mathrm{m,z,n}}$ exhibits spatial and temporal sparsity; therefore $\widetilde{\mathbf{d}}_{\mathrm{z,n}}$ is also a sparse signal. The system in equation is an underdetermined system due to the fact that $N_p < M_{ngf}L$ and cannot be reliably solved using the traditional channel estimation methods [39].

Moreover the equation can be rearranged into structured sparse form according to [39] as

$$\mathbf{y}_{\mathrm{z,n}} = \mathbf{A}_{\mathrm{n}} \widetilde{\mathbf{h}}_{\mathrm{z,n}} + \mathbf{w}_{\mathrm{z,n}} \tag{13}$$

where $\widetilde{\mathbf{h}}_{\mathrm{z,n}} = [\mathbf{h}_{1,\mathrm{z,n}}^{\mathrm{T}}, \mathbf{h}_{2,\mathrm{z,n}}^{\mathrm{T}}, \ldots, \mathbf{d}_{\mathrm{L,z,n}}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M_{ngf}L\times 1}$ is a structured sparse equivalent CIR vector; $\mathbf{h}_{\mathrm{l,z,n}} = [d_{1,\mathrm{z,n}}[l], d_{2,\mathrm{z,n}}[l], \ldots, d_{M_{ng},\mathrm{z,n}}[l]]^{\mathrm{T}}$ for $1 \le l \le L$ and after reformulation $\boldsymbol{\psi}_{\mathrm{n}}$ can be converted into $\mathbf{A}_{\mathrm{n}}$, where $\mathbf{A}_{\mathrm{n}}$ can be written as

$$\mathbf{A}_{\mathrm{n}} = [\mathbf{A}_{1,\mathrm{n}}, \mathbf{A}_{2,\mathrm{n}}, \ldots \mathbf{A}_{\mathrm{L,n}}] \in \mathbb{C}^{N_p\times M_{ngf}L} \tag{14}$$

where $\mathbf{A}_{l,n} = [\boldsymbol{\psi}_{1,n}^{(l)}, \boldsymbol{\psi}_{2,n}^{(l)}, \ldots, \boldsymbol{\psi}_{M_{ngf},n}^{(l)}] = [A_{1,n,l}, A_{2,n,l}, \ldots A_{M_{ngf},n,l}] \in \mathbb{C}^{Np\times M_{ngf}}$.

After converting to structured form, we will have N_g equations to be solved simultaneously corresponding to each subgroup. The received pilot vectors of $\mathrm{z_{th}}$ OFDM symbol for N_g subgroups can be expressed as follows.

$$\begin{aligned}
\mathbf{y}_{\mathrm{z},1} &= \mathbf{A}_1 \widetilde{\mathbf{h}}_{\mathrm{z},1} + \mathbf{w}_{\mathrm{z},1} \\
\mathbf{y}_{\mathrm{z},2} &= \mathbf{A}_2 \widetilde{\mathbf{h}}_{\mathrm{z},2} + \mathbf{w}_{\mathrm{z},2} \\
&\vdots \\
\mathbf{y}_{\mathrm{z,n}} &= \mathbf{A}_{\mathrm{n}} \widetilde{\mathbf{h}}_{\mathrm{z,n}} + \mathbf{w}_{\mathrm{z,n}} \\
&\vdots \\
\mathbf{y}_{\mathrm{z},N_g} &= \mathbf{A}_{N_g} \widetilde{\mathbf{h}}_{\mathrm{z},N_g} + \mathbf{w}_{\mathrm{z},N_g}
\end{aligned} \tag{15}$$

The system equation of each subgroup exhibits structured sparsity, which provides the motivation to apply CS theory to recover the high dimension structured sparse signal $\widetilde{\mathbf{h}}_{\mathrm{z},N_g}$ from the low dimension received pilot vector $\mathbf{y}_{\mathrm{z},N_g}$. The CS based sparse signal recovery/channel estimation will be explained in Section 5.

4.3. Proposed Pilot Design. Mostly wireless communication systems detect the data with the help of pilot signals. Specifically the purpose of pilot signals is to assist the receiver in estimating the wireless channels and then the receiver coherently detects the data on the basis of estimated channel [41].

In massive MIMO communication systems, due to large number of transmit antennas at BS, the number of channels gets prohibitively high and thus results in high pilot overhead for estimation of these channels. Therefore there is a need for methods to reduce the high pilot overhead in massive MIMO communication systems to achieve the target of high data rate.

In this research paper CS theory is applied to reduce the high pilot overhead based on the fact that the wireless channels undergo spatial and temporal sparsity. CS based pilot design significantly reduces the pilot overhead as compare to conventional pilot design. Figure 3 demonstrated the proposed, uniformly distributed, and identical pilot subcarrier for multiple antennas, while Figure 4 shows the convention pilot design (i.e., orthogonal pilots for different antennas). The proposed pilot design allows the system to provide more resources to data.

The CS based pilot design permits the antennas of each subgroup to occupy identical subcarriers for pilots within a group, while pilot subcarriers for each subgroup are completely different from each other as shown in Figure 3. The design is based on the choice of ξ_n. The guard interval can be calculated by $N_G = N/B\mathrm{x}(\text{maximum channel delay spread})$. Consider a set $O = \{1, 2, \ldots, N - N_G\}$. The ξ_n is a subset of O, represents the index of pilot subcarriers for n-th subgroup, and is identical for all the antennas within that group; the formula for creating the ξ_n can be given by

$$\begin{aligned}
\xi_n = \Bigg\{ & n + spacing * q : \left\lfloor \frac{M}{M_{ng}} \right\rfloor - \alpha \le N_g \\
& \le spacing,\ 0 < q < N_p - 1 \text{ and } sapcing \\
& = \left\lfloor \frac{N - N_G}{N_p} \right\rfloor \ge \left\lfloor \frac{M}{M_{ng}} \right\rfloor - \alpha \Bigg\}
\end{aligned} \tag{16}$$

where $N_p > L$ and by solving the inequality $\lfloor (N - N_G)/N_p \rfloor \ge \lfloor M/M_{ng} \rfloor - \alpha$ we have the following.

$$M_{ngf} \le N_p \le \frac{(N - N_G)}{\left(M - \alpha M_{ng}\right)} M_{ng} \tag{17}$$

There are N_G unique sets associated with ξ_n, i.e., $\xi_1 \ldots .. \xi_n \ldots \xi_{Ng}$. There will be total $N_g N_p$ subcarriers occupied

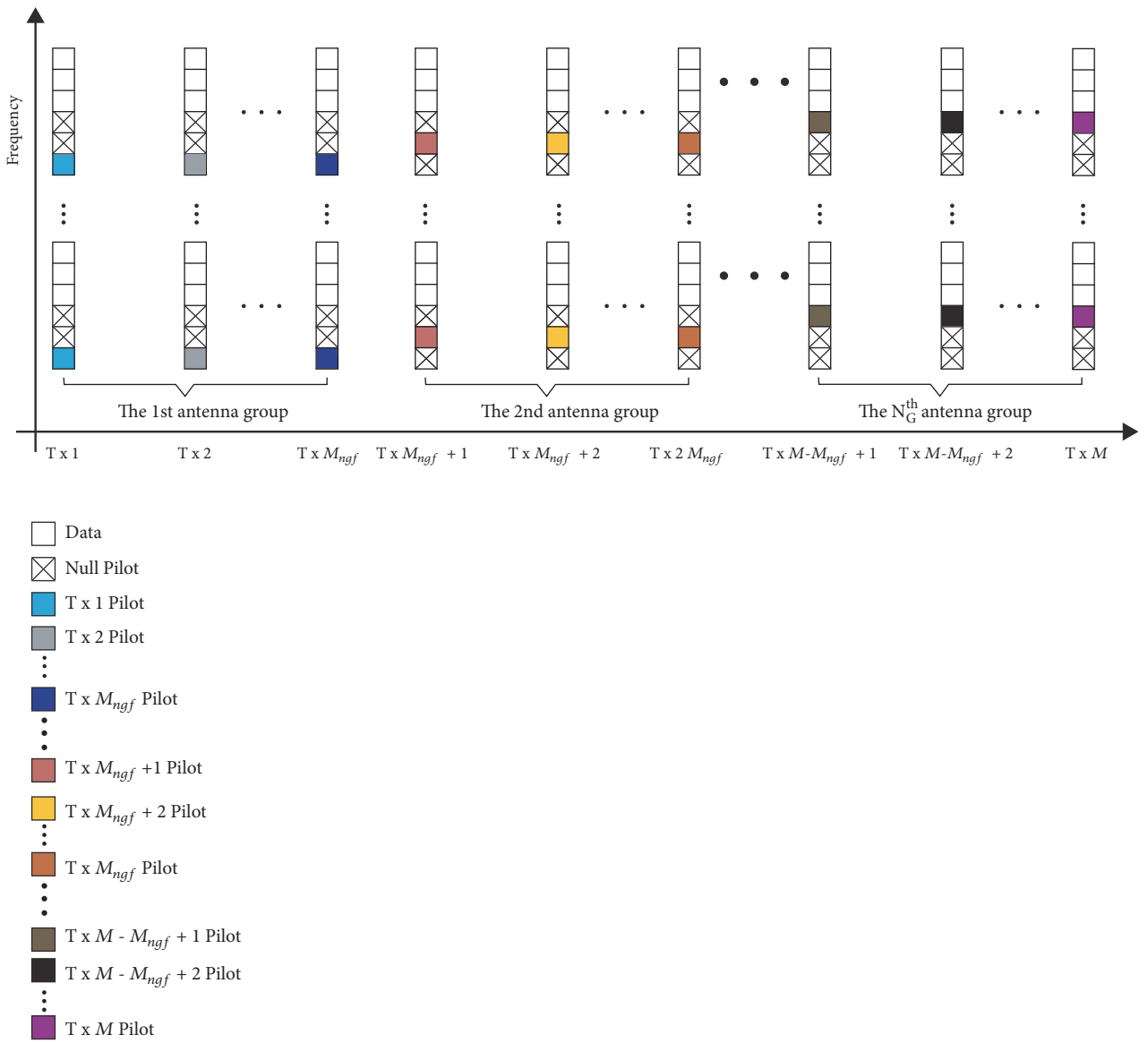


Figure 3: Proposed pilot design.

for pilot vectors for M transmit antennas at BS. Figure 3 demonstrates the proposed pilot design.

$N_p = |\xi_n|_C$ is the number of subcarriers for pilot vector in OFDM symbol of n_{th} subgroup.

Pilot subcarriers of n_{th} group are the null pilots for all the other $N_g - 1$ subgroups whereas $N - N_G - N_g N_p$ subcarriers are available for data.

Pilot overhead ratio is defined by $\beta_p = N_g N_p / N$.

Figure 4 shows the conventional pilots in which different pilots are allocated to different antennas resulting in very high pilot overhead.

5. Massive MIMO Channel Estimation Based on Compressive Sensing

The basic idea presented by CS theory is to recover a signal which is sparse in some domain from extremely small amount of nonadaptive linear measurements by applying convex optimization. In a different opinion, it relates the precise recovery of a sparse vector of high dimension by reducing its dimension. From another point of view, the problem can be considered as calculation of a signal's sparse coefficient with respect to an overcomplete system. The concept of compressed sensing was primarily applied for random sensing matrices, which allow for a reduced amount of nonadaptive, linear measurements. These days, the idea of compressed sensing has been generally replaced by sparse recovery.

In (13) it can be seen that $\widetilde{\mathbf{h}}_{z,n}$ demonstrates structured sparsity in delay domain, where $\widetilde{\mathbf{h}}_{z,n}$ is $S_{z,m,n}$-sparse vector due to

$$P_{z,m,n} = \operatorname{supp}\left\{\widetilde{\mathbf{h}}_{z,n}\right\} = \left\{l : \left|\widetilde{h}_{z,n}[l]\right| > 0\right\} \quad \text{with } 1 \le l \le L \tag{18}$$

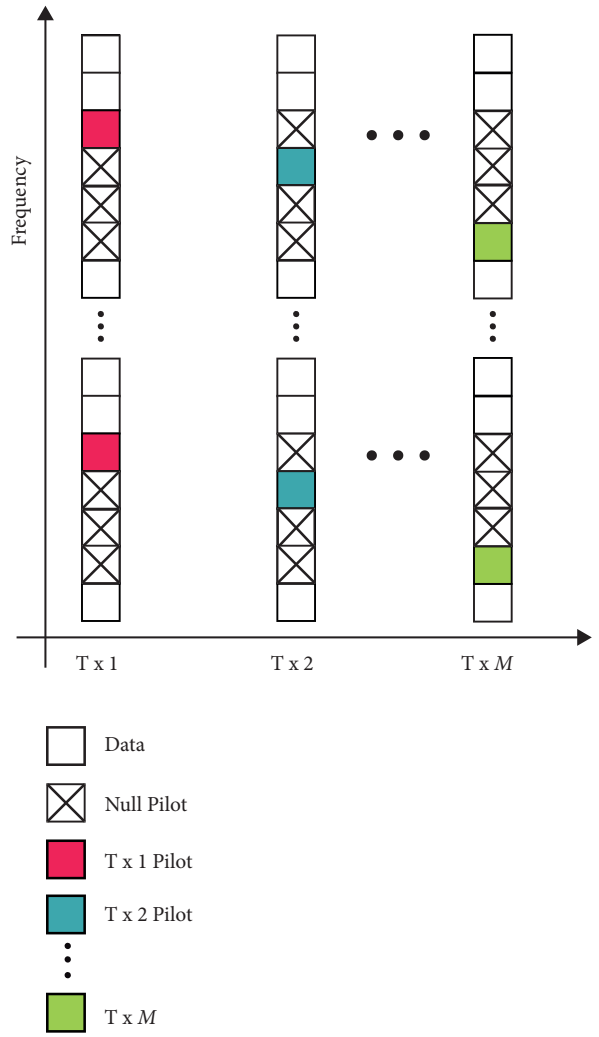


Figure 4: Conventional orthogonal pilot design.

where $S_{z,m,n} = |P_{z,m,n}|_C$ fulfilling $S_{z,m,n} \ll L$. And due to spatial sparsity we have the following.

$$P_{z,1,n} = P_{z,2,n} = \cdots = P_{z,M,n} \tag{19}$$

The desired small correlation of $\mathbf{A}_n$ according to CS theory is achieved which is described in [39]; therefore reliable sparse recovery is guaranteed. It is further shown in [39] that any two columns of $\mathbf{A}_n$ attain excellent cross correlation between them according to RMT since pilot design proposed in [39] is the special case of proposed pilot design. Therefore the proposed pilot design is simple to employ and also supportive in terms of compatibility with current wireless networks [39].

For K adjacent OFDM symbols having identical pattern of pilots, we have

$$\mathbf{Y}_n = \mathbf{A}_n \mathbf{H}_n + \mathbf{W}_n \tag{20}$$

where $\mathbf{Y}_n = [\mathbf{y}_{z,n}, \mathbf{y}_{z+1,n}, \ldots, \mathbf{y}_{z+K-1,n}] \mathbb{C}^{Np \times K}$.

The $\mathbf{A}_n$ derived in massive MIMO model satisfies the Structure Restricted Isometric Property (SRIP) condition. Specifically SRIP can be given by the following.

$$\sqrt{1-\delta} \left\| \mathbf{A}_{N_g} \right\|_F \leq \left\| \mathbf{A}_n \tilde{\mathbf{h}}_{z,n} \right\|_F \sqrt{1-\delta} \left\| \mathbf{A}_{N_g} \right\|_F \tag{21}$$

The definition and justification for SRIP are discussed in detail in [39], where $\delta \in [0, 1)$.

Our aim is to recover $\tilde{\mathbf{h}}_{z,n}$, given $\mathbf{y}_{z,n}$. Under the framework of CS theory, the massive MIMO channels $\tilde{\mathbf{h}}_{z,n}$ are estimated by means of the following problem:

$$\begin{aligned} \hat{\mathbf{h}} = \arg\min \quad & \left\| \tilde{\mathbf{h}}_{z,n} \right\|_1 \\ \text{s.t.} \quad & \left\| \mathbf{y}_{z,n} - \mathbf{A}_n \tilde{\mathbf{h}}_{z,n} \right\|_2 < \epsilon \end{aligned} \tag{22}$$

where ϵ is the noise variance. There are many algorithms that can solve the problem. For example, projected gradient methods or interior point methods can be used for applying convex optimization. A famous greedy algorithm is orthogonal matching pursuit (OMP).

Input: Sensing matrix $\mathbf{A}_n$ and noisy measurement vector $\mathbf{y}_{\mathrm{z,n}}$
Output: An sparse estimation $\widehat{\mathbf{h}}$ of channels$\{\mathbf{d}_{\mathrm{m,z,n}}\}_{m=1,\ n=1}^{\boldsymbol{m}=M_{ngf},\ n=N_g}$
Step 1 (Initialization)
1. $\check{\mathbf{h}}^0 \longleftarrow 0$
Trivial initial approximation
2. $\boldsymbol{v} \longleftarrow \mathbf{y}_{\mathrm{z,n}}$
Current samples = input samples
3. $k \longleftarrow 0$
Iterative index
4. $s \longleftarrow 1$
Initial sparsity level
Step 2 Solve the structure sparse vector $\widetilde{\mathbf{h}}_{\mathrm{z,n}}$ to (15)
Repeat
1. $k \longleftarrow k + 1$
2. $\boldsymbol{y} \longleftarrow \mathbf{A}_{\mathrm{n}}^{*}\mathbf{v}$
Make the signal proxy
3. $\Omega \longleftarrow \mathrm{supp}(\mathbf{y}_{2s})$
Identify large components
4. $\mathrm{T} \longleftarrow \Omega \cup \mathrm{supp}(\check{\mathbf{h}}^{k-1})$
Merge supports
5. $\check{\boldsymbol{h}}|_T \longleftarrow \boldsymbol{A}_{nT}^{\dagger}\mathbf{y}_{\mathrm{z,n}}$
Signal estimation by least squares
6. $\check{\boldsymbol{h}}|_{T^C} \longleftarrow 0$
7. $\check{\mathbf{h}}^k \longleftarrow \check{\boldsymbol{h}}_s$
Prune to get next approximation
8. $\boldsymbol{v} \longleftarrow \mathbf{y}_{\mathrm{z,n}} - \mathbf{A}_n\check{\mathbf{h}}^k$
Update the current samples
if $\|\boldsymbol{v}^k\|_2 < \|\boldsymbol{v}^{k+1}\|_2$
9. Iteration with fixed sparsity level
else
10. Update sparsity level $\check{\mathbf{h}}_s \longleftarrow \check{\mathbf{h}}^k$; $\boldsymbol{v}_s \longleftarrow \boldsymbol{v}^k$; $s \longleftarrow s + 1$
end if
Until stopping criterion true
Step 3 Obtain channels $\widehat{\boldsymbol{h}} = \check{\mathbf{h}}_{s-1}$ and obtain estimation of channels $\{\mathbf{d}_{\mathrm{m,z,n}}\}_{m=1,\ n=1}^{\boldsymbol{m}=M_{ngf},\ n=N_g}$ according to (11)-(15)

Algorithm 1: Proposed SUCoSaMPrecovery algorithm.

For channel estimation purpose we propose the SUCoSaMP algorithm derived from basic CoSaMP as described in Algorithm 1. There will be N_g similar parallel processing required for estimating the massive MIMO channels with N_g sub-antenna groups; i.e., the same algorithm will be working simultaneously with user to estimate channels of N_g subgroups.

There are many natural approaches of stopping the algorithm. We follow the following stopping criterion: if $\|\boldsymbol{v}^{k+1}\|_2 > \|\boldsymbol{v}_{s-1}\|_2$, the iteration is stopped [39]. The information of correct sparsity level $S_{z,m,n}$ is usually not available and also it is practically not possible to have prior knowledge of correct sparsity level, whereas information about sparsity level plays a significant role in compressive sensing problem of solving underdetermined system and it is also required as prior information by most of the CS based algorithms. The proposed SUCoSaMP algorithm does not require prior information of sparsity level because it adaptively acquires the sparsity level and avoids the unrealistic assumption of having prior information of correct sparsity level.

In steps 2.1 to 2.9 the target of SUCoSaMP algorithm is to obtain the solution $\widetilde{\mathbf{h}}_{\mathrm{z,n}}$ to (15) with fixed sparsity level s similar to conventional CoSaMP. The condition $\|\boldsymbol{v}^k\|_2 \leq \|\boldsymbol{v}^{k+1}\|_2$ shows that the solution $\widetilde{\mathbf{h}}_{\mathrm{z,n}}$ to (15) has been acquired and then the new iteration is started with updated sparsity level $s + 1$. This process is repeated until the stopping criteria are true and then the iteration is stopped. We get the solution to (15) with updated sparsity level and obtain the channels, i.e., $\widehat{\boldsymbol{h}} = \check{\mathbf{h}}_{s-1}$.

Note that the proposed SUCoSaMP algorithm works in the same way as the CoSaMP algorithm. The CoSaMP algorithm is basically based on basic OMP. The SUCoSaMP algorithm has incorporated some other ideas from the literature to present strong guarantees that OMP and CoSaMP cannot provide and to speed up the algorithm as compared to OMP.

The major advantage of SUCoSaMP over OMP, CoSaMP, and basic subspace pursuit (SP) algorithms is that it does not require prior knowledge of sparsity level which is an

Table 1: Major steps involved in CoSaMP and SUCoSaMP.

	CoSaMP	SUCoSaMP
1. Classification.	The algorithm creates a replacement of the residual from the existing samples and places the biggest components of the replacement.	Follows the same step of *CoSaMP*.
2. Support union.	The set of recently recognized components is joined with the set of components that emerge in the present approximation.	Follows the same step of *CoSaMP*.
3. Estimation.	The algorithm finds the solution of a least-squares problem to estimate the objective signal on the combined set of components.	Follows the same step of *CoSaMP*.
4. Pruning.	The algorithm generates a fresh estimation by keeping only the biggest entries in this least-squares approximation of signal.	Follows the same step of *CoSaMP*.
5. Sample Update.	Finally, the algorithm updates the samples such that they reflect the residual, the un-approximated elements of the signal.	Follows the same step of *CoSaMP*.
6. Stopping Criterion	Until stopping criterion is true.	Until first stopping criterion is true.
7. Sparsity level Update	None	The algorithm updates the sparsity level
8. Stopping Criterion	None	Until second stopping criterion is true.

Table 2: CoSaMP and SUCoSaMP hypothesis.

	CoSaMP	SUCoSaMP
1.	The sparsity level s is fixed. (i.e. initialization with fixed value of sparsity level).	The sparsity level s is not fixed (i.e. initialization with sparsity level equals 1). It adaptively acquires the correct sparsity level.
2.	The sensing matrix $\mathbf{A}_n$ has restricted isometry constant $\delta_{4s} \leq 0.1$.	$\mathbf{A}_n$ satisfies the Structure Restricted Isometric Property (SRIP) condition according to [39].
3.	The signal $\widetilde{\mathbf{h}}_{\mathrm{z,n}} \in \mathbb{C}^{M_{ngf}L\times 1}$ is random, except where prominent.	The signal $\widetilde{\mathbf{h}}_{\mathrm{z,n}} \in \mathbb{C}^{M_{ngf}L\times 1}$ is a structured sparse equivalent CIR vector.
4.	$\mathbf{w}_{\mathrm{z,n}}$ represents arbitrary noise vector	$\mathbf{w}_{\mathrm{z,n}}$ represents the Additive White Gaussian Noise of nth antenna group for zth OFDM symbol

unrealistic assumption, specifically in wireless communication scenario.

There are two differences between SUCoSaMP and CoSaMP. Firstly, SUCoSaMP estimates the channels; i.e., it recovers the high dimensional sparse vector by utilizing structured sparsity of massive MIMO channels from one vector of low dimension. Secondly, SUCoSaMP adaptively acquires the sparsity level. In contrast, the CoSaMP recovers the sparse vector without exploiting the structured sparsity and it requires prior information of correct sparsity level.

The proposed SUCoSaMP algorithm is exactly the same as CoSaMP algorithm up to step 2.9. The proposed algorithm stops the iteration with fixed sparsity level (i.e., the current value of s) if $\|\boldsymbol{v}^k\|_2 > \|\boldsymbol{v}^{k+1}\|_2$ and then it performs an additional step, that is, step # 2.10, to update the sparsity level by adding 1 to the current value of sparsity level (i.e., $s \longleftarrow s+1$).

There are two different types of iterations in SUCoSaMP, one on k and one on s, and finally the iterations on s are stopped if $\|\boldsymbol{v}^{k+1}\|_2 > \|\boldsymbol{v}_{s-1}\|_2$. Table 1 elaborates some further details of major steps involved in CoSaMP and SUCoSaMP algorithms. And Table 2 demonstrates the hypothesis of CoSaMP and SUCoSaMP algorithms.

Table 3: System parameters.

Parameter	Type/Value
Total number of transmit antennas	128
Number of transmit antennas in one sub-group	16
Number of antennas groups (N_g)	8
Modulation	16 QAM
Guard Interval	16
Number of pilot subcarriers (N_p)	32, 64
System bandwidth	20 MHz
DFT size	2048

6. Simulation Results

Simulations have been performed in MATLAB in order to verify the effectiveness of the proposed methods. Mean square error performance of proposed scheme is compared with the conventional OMP, CoSaMP, Structured Subspace Pursuit (SSP), and Adaptive Structured Subspace Pursuit (ASSP) algorithms. Simulation parameters are mentioned in Table 3 for the proposed system. The BS has 1D 1x128 antenna array ($M = 128$). The system bandwidth and

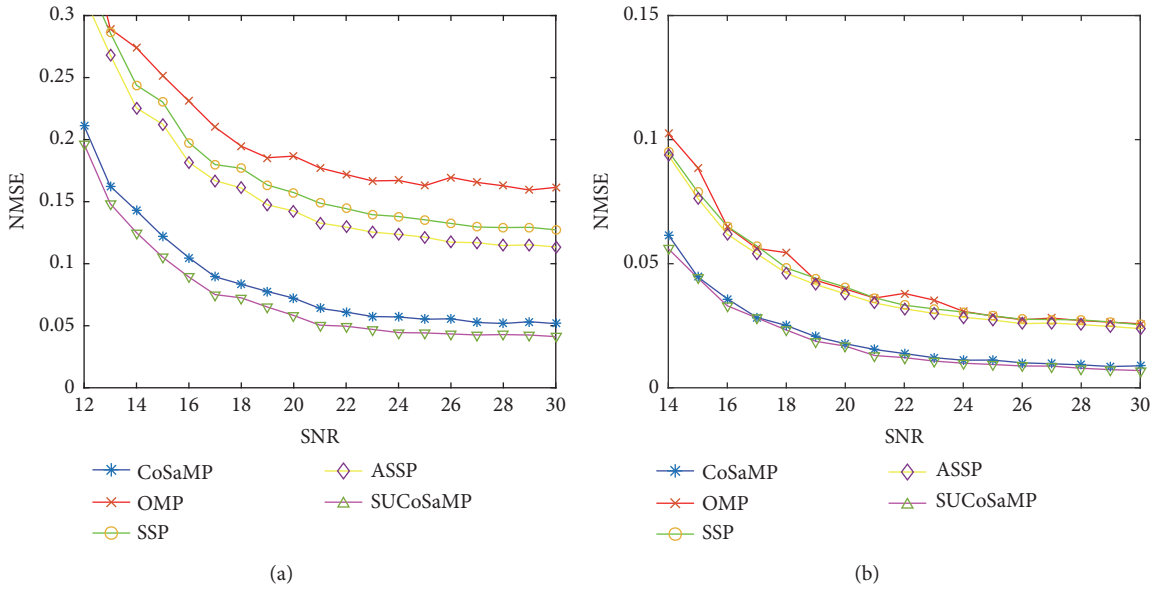


FIGURE 5: Comparison of MSE of each algorithm under different SNRs: (a) with $N_p = 32$, (b) with $N_p = 64$.

carrier frequency are set to $B = 20$ MHz and $f_c = 2$ GHz, respectively. There are $N_g = 8$ sub-antenna groups with 16 transmit antennas in each group to ensure the spatial channel sparsity within group. The OFDM subcarriers are set as $N = 2048$, guard interval is $N_G = 16$ which could fight the delay spread up to 6.4 μs, and 16QAM modulation is used. The numbers of pilot subcarrier N_p in OFDM symbol transmitted by each antenna in each antenna group and channel length L are varied over a reasonable range to verify the performance of the proposed system. The pilot positions are uniformly distributed according to (16) and are identical for the entire antennas within one group. The number of multipath channels is randomly chosen and the channel multipath amplitudes and positions follow Rayleigh and random distribution, respectively.

Figure 5 shows the MSE performance of the proposed SUCoSaMP algorithm with different values of N_p versus signal to noise ratio (SNR). The performance of SUCoSaMP is compared with the conventional algorithms OMP and CoSaMP and with SSP and ASSP. The sparsity level for OMP, CoSaMP, and SSP is known in simulations whereas the ASSP and proposed SUCoSaMP adaptively acquire the correct sparsity level. It is observed that the channel estimation performance of all the algorithms is improved by increasing pilot subcarriers N_p. And overall the proposed SUCoSaMP outperforms all the other algorithms. This is because the SUCoSaMP takes full advantage of structured sparsity of massive MIMO channels. Since the prior information of sparsity level of massive MIMO wireless channel is not a realistic assumption, the proposed SUCoSaMP algorithm has a clear advantage over the conventional algorithms. Furthermore the MSE gap between the proposed SUCoSaMP and the conventional algorithm remains almost constant as the SNR increases which makes SUCoSaMP superior in both low and high SNR wireless communication scenarios.

In Figures 6(a)–6(h) individual MSE performance versus SNR of different sub antenna groups (i.e., from antenna group N_1 to N_8) is presented for $N_p = 64$ and is compared with conventional algorithms OMP and CoSaMP and with SSP and ASSP. It can be seen in the figures that SUCoSaMP is evidently superior to conventional algorithms. This is because the provided sparsity level for conventional algorithms cannot be the actual sparsity of massive MIMO wireless channels. However, the adaptive process in ASSP and SUCoSaMP is realized by fixed increment in sparsity level; therefore the sparsity level estimation in these algorithms is slightly overestimated or underestimated. It is again observed that with high number of pilot subcarriers N_p, the performance of SUCoSaMP is improved along with the other algorithms. Furthermore it is observed that the MSE gap between the proposed SUCoSaMP and the conventional algorithm slightly increases or remains almost constant with the increase in SNR resulting in SUCoSaMP being better in different wireless communication environments with respect to SNR.

In Figures 7(a)–7(h) MSE performance versus SNR of individual antenna groups (i.e., from antenna group N_1 to N_8) is presented with $N_p = 32$ and comparison is provided with conventional algorithms OMP and CoSaMP and with SSP and ASSP. It can be seen from the figures that the SUCoSaMP algorithm for each antenna group can achieve huge MSE gains over conventional algorithms. However it is observed that the MSE performance of all algorithms slightly degraded with decrease in number of pilot subcarriers N_p. Results show that SUCoSaMP efficiently considers the effect of SNR and performs better in both high and low SNR scenarios.

7. Conclusion

This paper proposes a nonorthogonal pilot design and a CS based algorithm SUCoSaMP to estimate the massive

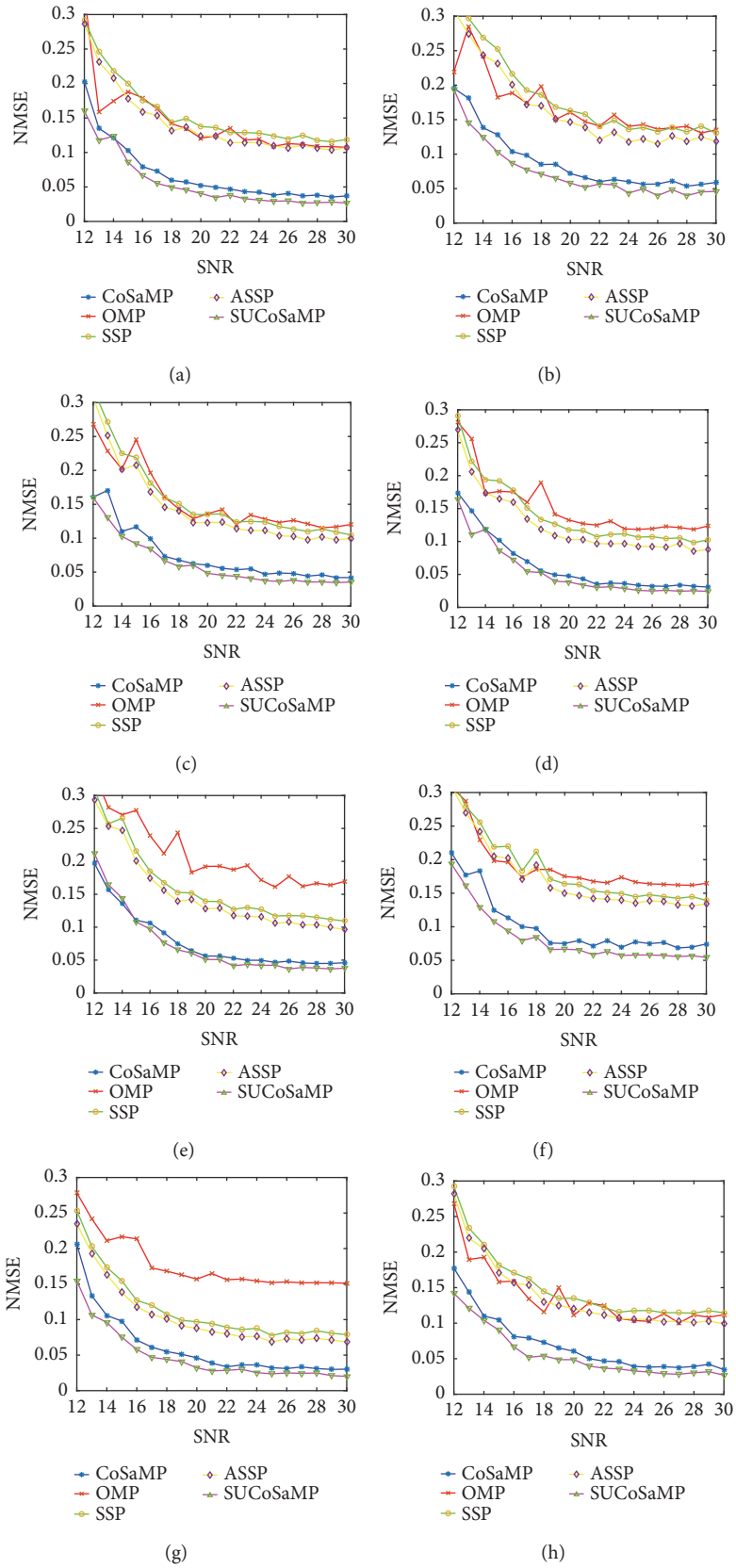


FIGURE 6: Comparison of MSE of each algorithm under different SNRs for individual antenna groups, with $N_p = 64$. (a) MSE performance versus SNR of 1st antenna group (N_1). (b) MSE performance versus SNR of 2nd antenna group (N_2). (c) MSE performance versus SNR of 3rd antenna group (N_3). (d) MSE performance versus SNR of 4th antenna group (N_4). (e) MSE performance versus SNR of 5th antenna group (N_5). (f) MSE performance versus SNR of 6th antenna group (N_6). (g) MSE performance versus SNR of 7th antenna group (N_7). (h) MSE performance versus SNR of 8th antenna group (N_8).

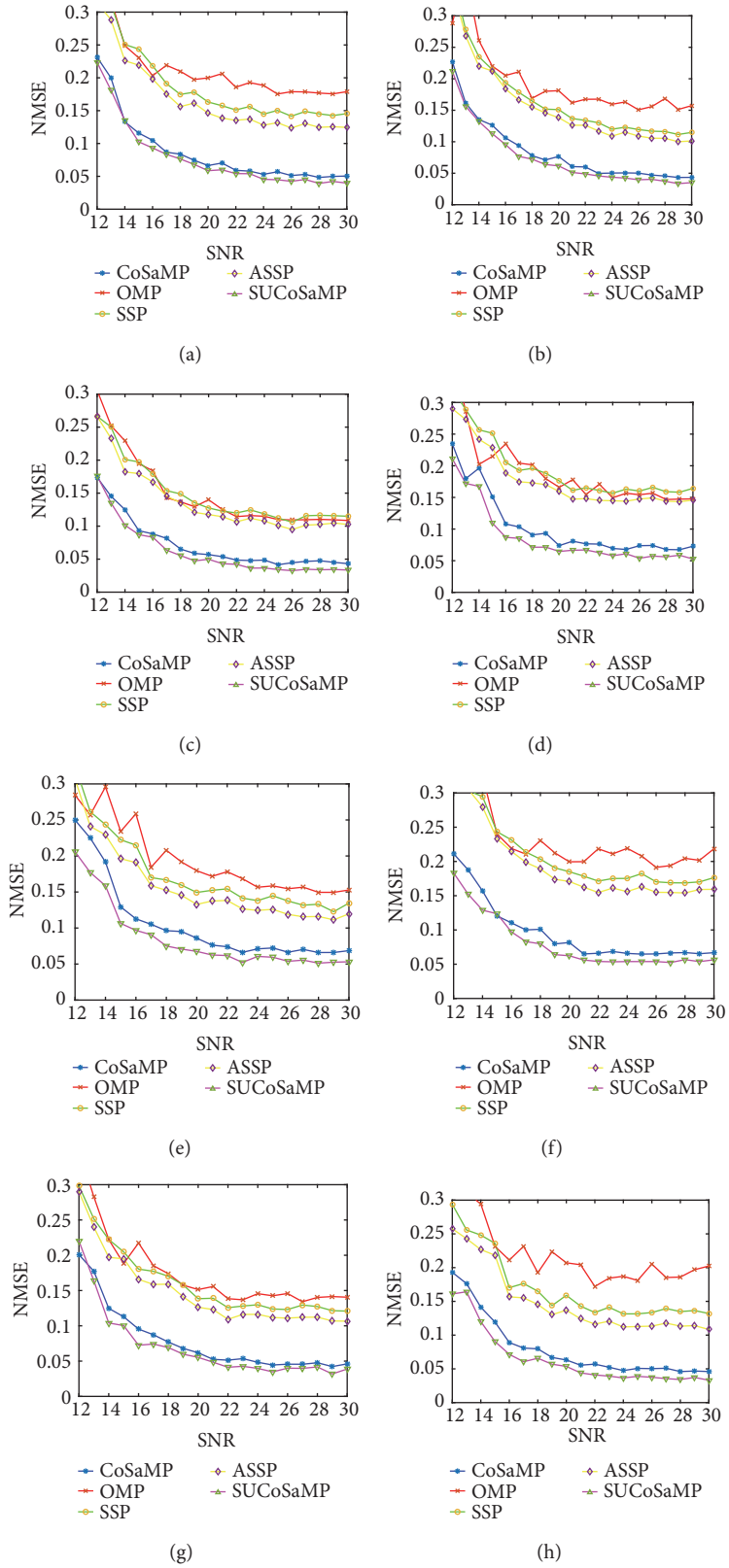


FIGURE 7: Comparison of MSE of each algorithm under different SNRs for individual antenna groups, with $N_p = 32$. (a) MSE performance versus SNR of 1st antenna group (N_1). (b) MSE performance versus SNR of 2nd antenna group (N_2). (c) MSE performance versus SNR of 3rd antenna group (N_3). (d) MSE performance versus SNR of 4th antenna group (N_4). (e) MSE performance versus SNR of 5th antenna group (N_5). (f) MSE performance versus SNR of 6th antenna group (N_6). (g) MSE performance versus SNR of 7th antenna group (N_7). (h) MSE performance versus SNR of 8th antenna group (N_8).

MIMO channels. The reduction of high pilot overhead in massive MIMO systems and the recovery ability when the sparsity level of massive MIMO channels is unknown are the main focus of this research. By taking advantage of spatial and temporal common sparsity of massive MIMO channels in delay domain, the proposed nonorthogonal pilot design and channel estimation scheme under the frame work of CS theory significantly reduce the pilot overheads for massive MIMO systems and also outperform the conventional algorithms in performance. This research may be extended by incorporating the proposed schemes in the multicell scenarios.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] C.-X. Wang, F. Haider, X. Gao et al., "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.

[2] F. Rusek, D. Persson, and B. K. Lau, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal*, 2013.

[3] H. Yang, Y. Fan, D. Liu, and Z. Zheng, "Compressive sensing and prior support based adaptive channel estimation in massive MIMO," in *Proceedings of the 2nd IEEE International Conference on Computer and Communications, ICCC '16*, IEEE, 2016.

[4] J. G. Andrews, S. Buzzi, and W. Choi, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[5] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications Magazine*, vol. 21, no. 2, pp. 106–112, 2014.

[6] W. Xu, T. Shen, Y. Tian, and Y. Wang, "Compressive channel estimation exploiting block sparsity in multi-user massive MIMO systems," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference, WCNC '17*, IEEE, 2017.

[7] J. Zhang, B. Zhang, S. Chen, and X. Mu, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE Journal*, 2014.

[8] E. Bjonson, J. Hoydis, and M. Kountouris, "Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits," *IEEE Transactions*, 2014.

[9] Y. S. Cho, J. Kim, W. Yang, and C. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, 2010.

[10] Y. Xu, G. Yue, and S. Mao, "User grouping for massive MIMO in FDD systems: new design methods and analysis," *IEEE Access*, vol. 2, pp. 947–959, 2013.

[11] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[12] L. Correia, *Mobile Broadband Multimedia Networks: Techniques, Models and Tools for 4G*, 2010.

[13] Z. Zhou, X. Chen, and D. Guo, "Sparse channel estimation for massive MIMO with 1-bit feedback per dimension," in *Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC '17*, 2017.

[14] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Transactions on Signal Processing*, 2010.

[15] X. Ma, F. Yang, S. Liu, and J. Song, "Training sequence design and optimization for structured compressive sensing based channel estimation in massive MIMO systems," in *Proceedings of the Globecom Work (GC '16)*, 2016.

[16] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 6, pp. 1615–1624, 2003.

[17] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Transactions on Wireless Communications*, 2006.

[18] Y.-H. Nam, Y. Akimoto, Y. Kim, M.-I. Lee, K. Bhattad, and A. Ekpenyong, "Evolution of reference signals for LTE-advanced systems," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 132–138, 2012.

[19] L. Lu, G. Y. Li, and A. L. Swindlehurst, "An overview of massive MIMO: benefits and challenges," *IEEE Journal*, 2014.

[20] N. Gonzalez-Prelcic, K. T. Truongt, C. Rusu, and R. W. Heath, "Compressive channel estimation in FDD multi-cell massive MIMO systems with arbitrary arrays," in *Proceedings of the 2016 IEEE Globecom Workshops, GC Wkshps '16*, 2016.

[21] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 251–263, 2013.

[22] M. S. Sim, J. Park, and C.-B. Chae, "Compressed channel feedback for correlated massive MIMO," *Commun.*, 2016.

[23] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *IEEE Electronics Letters*, vol. 50, no. 12, pp. 896–898, 2014.

[24] C. Qi and L. Wu, "Uplink channel estimation for massive MIMO systems exploring joint channel sparsity," *IEEE Electronics Letters*, vol. 50, no. 23, pp. 1770–1772, 2014.

[25] Z. Gao, L. Dai, Z. Lu, and C. Yuen, "Super-resolution sparse MIMO-OFDM channel estimation based on spatial and temporal correlations," *IEEE Communications Letters*, vol. 18, no. 7, pp. 1266–1269, 2014.

[26] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proceedings of the 2013 IEEE Wireless Communications and Networking Conference, WCNC '13*, 2013.

[27] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, 2015.

[28] W. Shen, L. Dai, B. Shim, and S. Mumtaz, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive

MIMO systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2178–2181, 2015.

[29] X. Ma, F. Yang, S. Liu, and J. Song, "Doubly selective channel estimation for MIMO systems based on structured compressive sensing," in *Proceedings of the 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC '17)*, 2017.

[30] Y. Ding and B. D. Rao, "Dictionary learning based sparse channel representation and estimation for FDD massive MIMO systems," 2016.

[31] B. Gong, L. Gui, Q. Qin, and X. Ren, "Block distributed compressive sensing-based doubly selective channel estimation and pilot design for large-scale MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9149–9161, 2017.

[32] S. Sun and T. S. Rappaport, "Millimeter Wave MIMO channel estimation based on adaptive compressed sensing," in *Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops '17)*, 2017.

[33] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, 2014.

[34] D. Hu, X. Wang, and L. He, "A new sparse channel estimation and tracking method for time-varying OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4648–4653, 2013.

[35] T. Santos, J. Karedal, and P. Almers, "Modeling the ultra-wideband outdoor channel: measurements and parameter extraction method," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 282–290, 2010.

[36] I. E. Telatar and D. N. Tse, "Capacity and mutual information of wideband multipath fading channels," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 46, no. 4, pp. 1384–1400, 2000.

[37] I. Khan, M. Singh, and D. Singh, "Compressive sensing-based sparsity adaptive channel estimation for 5G massive MIMO systems," *Applied Sciences*, vol. 8, no. 5, p. 754, 2018.

[38] Y. Nan, L. Zhang, and X. Sun, "Weighted compressive sensing based uplink channel estimation for TDD massive MIMO sytems," *IET Communications*, vol. 11, no. 3, pp. 355–361, 2017.

[39] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 601–617, 2016.

[40] J. Choi, B. Shim, Y. Ding, B. Rao, and D. Kim, "Compressed sensing for wireless communications: useful tips and tricks," *IEEE Communications Survey and Tutorials*, vol. 19, no. 3, pp. 1527–1550, 2017.

[41] A. Lozano and N. Jindal, "Optimum pilot overhead in wireless communication: a unified treatment of continuous and block-fading channels," in *Proceedings of the 2010 European Wireless Conference (EW '10)*, 2010.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 3137927, 12 pages
https://doi.org/10.1155/2019/3137927

Research Article

Generalized Complex Quadrature Spatial Modulation

Manar Mohaisen

Department of EEC Engineering, Korea Tech, Cheonan, Republic of Korea

Correspondence should be addressed to Manar Mohaisen; manar.subhi@koreatech.ac.kr

Received 7 September 2018; Revised 6 February 2019; Accepted 24 February 2019; Published 28 April 2019

Guest Editor: Mohamad K. A. Rahim

Spatial modulation (SM) is a multiple-input multiple-output (MIMO) system that achieves a MIMO high spectral efficiency while maintaining the transmitter computational complexity and requirements as low as those of the single-input systems. The complex quadrature spatial modulation (CQSM) builds on the QSM scheme and improves the spectral efficiency by transmitting two signal symbols at each channel use. In this paper, we propose two generalizations of CQSM, namely, generalized CQSM with unique combinations (GCQSM-UC) and with permuted combinations (GCQSM-PC). These two generalizations perform close to CQSM or outperform it, depending on the system parameters. Also, the proposed schemes require much less transmit antennas to achieve the same spectral efficiency of CQSM, for instance, assuming 16-QAM, GCQSM-PC, and GCQSM-UC require 10 and 15 transmit antennas, respectively, to achieve the same spectral of CQSM which is equipped with 32 antennas.

1. Introduction

Multiple-input multiple-output (MIMO) techniques are capable of satisfying the continuous demand in high data rates in communication systems. Index modulation (IM) is a group of MIMO techniques that use the indices of a given resource(s) of the communication system to convey additional information, besides that carried by the signal symbols [1, 2]. These indices represent distinct antennas [3], spreading codes [4], polarities [5], subcarriers [6], and combinations of angles [7], among others. A virtual spatial modulation (VSM) was also proposed in [8], where index modulation is performed on the parallel channels resulting from the singular value decomposition of the MIMO channel matrix.

Spatial modulation (SM) has attracted an increasing interest from researchers due to its low-complexity and high data rates. In SM, information is transmitted through the signal symbols, drawn from any conventional constellation sets, and the index of the antenna from which the signal symbol is transmitted [3]. To achieve a high spectral efficiency, SM requires relatively high number of transmit antennas. A parallel SM, denoted by half-full transmit-diversity SM (HFTD-SM), is proposed recently, where antennas are grouped into two-antenna groups. The conventional SM is then performed using the same signal symbol in each group [9, 10].

A conventional generalized SM (GSM) transmits the same signal symbol from a combination of antennas, rather than one as in SM, leading to a reduction in the number of transmit antennas required to achieve a given spectral efficiency [11]. A different implementation of the GSM was proposed in [12], referred to as MA-SM, in which each activated antenna in a given combination transmits a different signal symbol. This improves the spectral efficiency at the cost of additional hardware requirements and computational complexity. GSM was extended to the multiuser scenario in [13].

Quadrature spatial modulation (QSM) is another technique that extends the spatial constellation into in-phase and quadrature dimensions that are used to transmit the real and imaginary part, respectively, of a single signal symbol [14]. In [15], a generalized QSM (GQSM) that combines the benefits of spatial multiplexing and QSM was proposed. Antennas are grouped and independent QSM modulation is performed in each group. The hardware design of the generalized quadrature space shift keying modulation (GQSSK) and generalized QSM (GQSM) using a single RF chain was investigated in [16].

Building on the QSM, a complex QSM (CQSM) improves the spectral efficiency by transmitting two signal symbols at each channel use from antennas whose indices also carry information [17]. When both signal symbols are transmitted from the same antenna, the resulting combination is drawn from the Minkowski sum of the two sets from which the symbols are drawn. The second modulation set is designed as a rotated version of the first, where the rotation angle is optimized to minimize the average bit-error rate. To avoid transmitting the two signal symbols from the same antenna, the transmitter is equipped with an additional antenna that is used to transmit the second symbol only when both symbols are supposed to be transmitted from the same antenna in CQSM. The proposed modification is named improved CQSM (ICQSM) in [18]. The modulation set design of the ICQSM was addressed in [19]. Notice that the CQSM can be conceptualized based on the conventional SM or the GSM, and not as an extension of the QSM. While this approach is correct, we built CQSM as an extension of the QSM where instead of transmitting the real and imaginary parts of a single signal symbol from the in-phase and quadrature spatial dimensions, two complex-valued signal symbols are transmitted through the two spatial dimensions. To be consistent with our previous work, we keep the same name that explicitly includes QSM.

Contributions. The contributions of this paper are summarized as follows:

(i) We propose two generalized CQSM schemes. The first is GCQSM with unique combinations (GCQSM-UC), where the spatial symbols are generated as in generalized GSM. Each resulting combination is then split into two subsets. The first subset of antennas is used to transmit the first symbol; the second subset is used to transmit the second signal symbol. The second scheme is named GCQSM with permuted combinations (GCQSM-PC). In GCQSM-PC the antenna combinations from which the first symbol is transmitted are generated as the full set of combinations of a given length. For each combination associated with the first symbol, the corresponding list of combinations for the second symbol is generated such that no overlap occurs between antennas used for transmitting the two symbols. This algorithm reduces the number of transmit antennas required to achieve a given spectral efficiency.

(ii) We also propose an analytical method to optimize the rotation angle for CQSM and ICQSM systems. The obtained rotation angle minimizes the upper-bound of pairwise error probability.

Based on the simulation results, the proposed schemes perform very close to CQSM, while requiring much smaller number of antennas to achieve the same spectral efficiency. For instance, CQSM, GCQSM-UC, and GCQSM-PC require 32, 15, and 10 transmit antennas to achieve a spectral efficiency of 14 bits per channel use (bpcu), assuming 2 bits per signal symbol.

Notations. We denote the spectral efficiency achieved by spatial and signal symbols ES_{spa} and ES_{sig}, respectively, and the total spectral efficiency $\mathrm{ES} = \mathrm{ES}_{sig} + \mathrm{ES}_{spa}$. The vector $\mathbf{e}_a^b$ is the ath column of the $b \times b$ identity matrix. $A = \{a_1, \ldots, a_K\}$ is a set, where the order of elements does not matter, and $B = (b_1, \ldots, b_K)$ is a tuple where order matters. In the following, $(\cdot)^T$ and $(\cdot)^H$ denote the matrix/vector transpose and Hermitian transpose, respectively. $Q(\cdot)$ is the Gaussian tail function, or simply the Q-function. A signal symbol is a complex element drawn from a quadrature amplitude modulation (QAM) or phase shift keying (PSK) set. The number of bits carried by each signal symbol is equal to q, where 2^q is the cardinality of the modulation set. A spatial symbol is the index(es) of the antenna(s) from which a single or several signal symbols are transmitted.

The rest of this paper is organized as follows. In Section 2, we describe the system model and review related works. In Section 3, we introduced the proposed generalized CQSM techniques and analyze their error performance in Section 4. In Section 5, we formulate the search of the optimal rotation angle for CQSM and ICQSM as an optimization problem that reduced the asymptotic upper-bound on the error probability. The optimization of the rotation angle of the proposed generalized schemes is addressed in Section 6. The simulation results are given in Section 7 and conclusions are drawn in Section 8.

2. System Model and Related Works

2.1. System Model. Consider a MIMO system with n_T transmit and n_R receive antennas. The system equation is given as follows.

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ and $\mathbf{n}$ are the channel matrix and the noise vector whose elements are i.i.d. centered circularly symmetric complex Gaussian and have a variance of one and σ_n^2, respectively, and $\mathbf{s}$ is the transmitted vector. In SM techniques, the vector $\mathbf{s}$ contains a few nonzero elements.

2.2. Spatial Modulation. In SM, both a signal symbol and the index of the antenna from which it is transmitted carry information. As such, SM achieves an improved spectral efficiency while keeping the transmitter as simple as that of the single-input communication systems. Accordingly, the SM system is modeled as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{e}_i^{n_T} s_k + \mathbf{n} = \mathbf{h}_i s_k + \mathbf{n}. \tag{2}$$

The spectral efficiency, which is equivalent to the capacity at high signal-to-noise ratio (SNR), is equal to $q + N$ bits per channel use (bpcu), with $N = \log_2(n_T)$.

2.3. Generalized Spatial Modulation. In SM, the number of transmit antennas increases exponentially in terms of the number of bits carried by each spatial symbol. That is, $n_T = 2^N$, where N is the number of bits per spatial symbol. GSM

reduces the number of transmit antennas by sending the same signal symbol from a combination of two or more antennas. Let n_U denote the number of active antennas at each channel use; then the number of spatial symbols, also referred to as combinations, for a given n_T is $\binom{n_T}{n_U}$. Since the number of spatial symbols should be a power of two, only $2^{\mathrm{ES}_{spa}}$ combinations can be used, where $\mathrm{ES}_{spa} = \lfloor \log_2 \binom{n_T}{n_U} \rfloor$. Assuming $n_U = 2$, GSM requires 7 transmit antennas to achieve $\mathrm{ES}_{spa} = 4$ bits per spatial symbol, whereas SM requires 16 antennas to achieve the same spectral efficiency.

Let $\boldsymbol{l} = \{l_1, \ldots, l_{n_U}\} \in \mathbf{L}$ be a spatial symbol, where $\mathbf{L}$ is the set of spatial symbols that can be used for transmission. Then the received vector in the case of GSM is given by

$$\mathbf{y} = \frac{s_k}{\sqrt{n_U}} \sum_{i \in \boldsymbol{l}} \mathbf{h}_i + \mathbf{n}. \tag{3}$$

2.4. Multi-Active Spatial Modulation. In contrast to GSM, in which a single signal symbol is transmitted from a combination of antennas, MA-SM transmits a different signal symbol from each activated antenna. Let n_U be the number of activated antennas; then the received vector is given by

$$\mathbf{y} = \sum_{i \in \boldsymbol{l}} \mathbf{h}_{l_i} s_i + \mathbf{n}, \tag{4}$$

where $\boldsymbol{l} = \{l_1, \ldots, l_{n_U}\} \in \mathbf{L}$, and $\mathbf{L}$ is the list of spatial symbols of MA-SM, which is equivalent to that of GSM. Additionally, a 3-dimensional constellation set is proposed in which each spatial symbol is associated with a unique rotation angle. Signal symbols transmitted from a given spatial symbol are rotated before transmission. According to this description, MA-SM benefits from the moderate computational complexity and low hardware requirements of SM and the high multiplexing gain of the vertical-Bell Labs layered space-time (V-BLAST) system. The spectral efficiency of MA-SM is equal to $(\mathrm{ES}_{spa} + q \times n_U)$, where ES_{spa} is defined in Section 2.3.

2.5. Complex Quadrature Spatial Modulation. QSM expands the spatial constellation into in-phase and quadrature dimensions. At each channel use, a single signal symbol is transmitted: the real part of the signal symbol is transmitted through the in-phase spatial dimension; the imaginary through the quadrature dimension. As such, the spectral efficiency of QSM is $q + 2\log_2(n_T)$, where $\mathrm{ES}_{spa} = 2\log_2(n_T)$ and $\mathrm{ES}_{sig} = q$.

CQSM transmits two signal symbols at each channel use, leading to a spectral efficiency of $2q + 2\log_2(n_T)$, where $\mathrm{ES}_{spa} = 2\log_2(n_T)$ and $\mathrm{ES}_{sig} = 2q$. The first signal symbol $s_{k_1} \in \Omega_a$ is transmitted from the i_1th transmit antenna, and the second signal symbol $s_{k_2'} \in \Omega_b$ is transmitted from the i_2th antenna. To make the distinction between s_{k_1} and $s_{k_2'}$ at the receiver side, Ω_b is obtained as follows:

$$\Omega_b = \left\{ s_{k_2'} \mid s_{k_2'} = s_{k_1} e^{j\theta}, s_{k_1} \in \Omega_a \right\}. \tag{5}$$

The rotation angle θ is optimized such that the probability of error is reduced. The received vector of the CQSM is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} = \mathbf{h}_{i_1} s_{k_1} + \mathbf{h}_{i_2} s_{k_2'} + \mathbf{n}, \tag{6}$$

which is expanded to

$$\mathbf{y} = \begin{cases} \mathbf{h}_{i_1} s_{k_1} + \mathbf{h}_{i_2} s'_{k_2} + \mathbf{n} & \text{if } i_1 \neq i_2 \\ \mathbf{h}_i \left(s_{k_1} + s'_{k_2} \right) = \mathbf{h}_i s_{k_3''}, & \text{if } i_1 = i_2 = i \end{cases} \tag{7}$$

where $s_{k_3''} \in \Omega_c$ and $\Omega_c = \Omega_a \oplus \Omega_b$, with $\oplus$ denoting the Minkowski sum [20]. The full modulation set $\Omega_d = \Omega_a \cup \Omega_b \cup \Omega_c$ is of size $(|\Omega_a| + |\Omega_b| + |\Omega_a| \times |\Omega_b|)$. For instance, assuming $|\Omega_a| = |\Omega_b| = 16$, the number of symbols in Ω_d becomes 288. This high density of symbols reduces the Euclidean distance among signal symbols, leading to degradation in the error performance.

2.6. Improved Complex Quadrature Spatial Modulation. In ICQSM, the transmitter is equipped with an additional antenna, indexed n_K with $n_K = n_T + 1$, which is used only when $i_1 = i_2$. In this case, the first signal symbol is transmitted from its designated antenna and the second symbol is transmitted from the additional antenna. Accordingly, the ICQSM system is modeled as follows:

$$\mathbf{y} = \begin{cases} s_{k_1} \mathbf{h}_{i_1} + s_{k_2'} \mathbf{h}_{i_2} + \mathbf{n}, & \text{if } i_1 \neq i_2 \\ s_{k_1} \mathbf{h}_{i_1} + s_{k_2'} \mathbf{h}_{n_K} + \mathbf{n}, & \text{if } i_1 = i_2 \end{cases} \tag{8}$$

This strategy reduces the size of the total modulation set Ω_d from $(|\Omega_a| + |\Omega_b| + |\Omega_a| \times |\Omega_b|)$ in the case of CQSM to only $(|\Omega_a| + |\Omega_b|)$ in ICQSM. This increases the Euclidean distance among the signal symbols and hence improves the error performance.

Based on the above descriptions, CQSM activates either one or two antennas, whereas ICQSM always has two active antennas. To perform a fair comparison between these two systems and MA-SM, we assume that $n_U = 2$. That is, the hardware requirements of the transmitter and the computational complexity of the detection algorithm are equivalent. Since both systems transmit two signal symbols at each channel use, the comparison is conducted in terms of the spectral efficiency achieved by spatial symbols. SM, on the other hand, transmits a single signal symbol at each channel use. Table 1 lists the number of transmit antennas required to achieve a given number of bits per spatial symbol (SE_{spa}) by several systems. For instance, CQSM and ICQSM require 3 and 2 antennas less than MA-SM to achieve the same SE_{spa} of 4 bpcu. This gap increases for higher SE_{spa}.

3. Generalized Complex Quadrature Spatial Modulation

Both CQSM and ICQSM require the same number of RF chains, and ICQSM requires one more physical transmit antenna compared to CQSM. Transmitting each signal symbol from a combination of antennas, instead of a single antenna, reduces the number of transmit antennas required to achieve a given SE_{spa}. In the following subsections, we introduce two generalizations of CQSM, namely, generalized CQSM with unique combinations (GCQSM-UC) and generalized CQSM with permuted combinations (GCQSM-PC).

Table 1: The number of transmit antennas required by several systems to achieve the same number of bits per spatial symbol (SE_{spa}).

SE_{spa}	SM	MA-SM	ICQSM	CQSM
2	4	4	3	2
4	16	7	5	4
6	64	12	9	8

Table 2: An example of the spatial symbols $(\boldsymbol{i}_1, \boldsymbol{i}_2)$ that can be used for transmission by GCQSM-UC, assuming $n_T = 6$ and $n_U = 2$.

$\boldsymbol{i}_1$	$\boldsymbol{i}_2$
{1, 2}	{3, 4}
{1, 2}	{3, 5}
{1, 2}	{3, 6}
{1, 2}	{4, 5}
{1, 2}	{4, 6}
{1, 2}	{5, 6}
{1, 3}	{4, 5}
{1, 3}	{4, 6}
{1, 3}	{5, 6}
{1, 4}	{5, 6}
{2, 3}	{4, 5}
{2, 3}	{4, 6}
{2, 3}	{5, 6}
{2, 4}	{5, 6}
{3, 4}	{5, 6}

In the following, we denote by $\boldsymbol{i}_1$ and $\boldsymbol{i}_2$ the set of antennas used for transmitting the first and second signal symbols, respectively. In the sequel, the tuple $A = (B, C)$ is composed of the two sets B and C, where the order of the sets in the tuple matters but the order of the elements in each set does not.

3.1. Generalized CQSM with Unique Combinations. Let n_U be the number of antennas from which each of the two signal symbols is transmitted; then for a given n_T, the number of combinations that can be used for transmission is $\binom{n_T}{2n_U}$. Assuming that $\boldsymbol{i} = \{\boldsymbol{i}_1, \boldsymbol{i}_2\}$ is a spatial symbol, i.e., a combination of antennas of length $2n_U$; then the first and second n_U antennas are used to transmit the first and second signal symbols, respectively. Table 2 depicts an example of the spatial symbols for $n_T = 6$ and $n_U = 2$. According to this description, there will be no overlap between the sets of antennas used to transmit the first and second signal symbols. Since the number of spatial symbols should be a power of two, then only $2^{SE_{spa}}$ signal symbols can be used for transmission, where $SE_{spa} = \left\lfloor \log_2 \binom{n_T}{2n_U} \right\rfloor$.

Let $s_{k_1} \in \Omega_a$ and $s_{k_2'} \in \Omega_b$ be the two signal symbols to be transmitted at a given channel use; then the received vector is given by

$$\mathbf{y} = \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} \mathbf{h}_i + \frac{s_{k_2'}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} \mathbf{h}_i + \mathbf{n}. \tag{9}$$

The receiver employs the maximum-likelihood principle to recover the two signal symbols and the spatial symbol as follows:

$$\begin{aligned} \left(\boldsymbol{i}^*, s_{k_1}^*, s_{k_2'}^*\right) &= \underset{\substack{s_{k_1} \in \Omega_a, s_{k_2'} \in \Omega_b \\ \boldsymbol{i} = (\boldsymbol{i}_1, \boldsymbol{i}_2) \in \mathbb{L}}}{\arg\min} \left\| \mathbf{y} - \mathbf{g} \right\|^2 \\ &= \underset{\substack{s_{k_1} \in \Omega_a, s_{k_2'} \in \Omega_b \\ \boldsymbol{i} = (\boldsymbol{i}_1, \boldsymbol{i}_2) \in \mathbb{L}}}{\arg\min} \left\| \mathbf{g} \right\|^2 - 2\Re\left(\mathbf{y}^H \mathbf{g}\right), \end{aligned} \tag{10}$$

where

$$\mathbf{g} = \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} \mathbf{h}_i + \frac{s_{k_2'}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} \mathbf{h}_i \tag{11}$$

GCQSM-UC can be easily extended to transmitting more than two signal symbols per channel use at the cost of requiring more RF chains at the transmitter side. This is possible because the error performance of the system does not depend on the rotation angle θ. This will be addressed in more detail in subsequent sections.

3.2. Generalized CQSM with Permuted Combinations. To reduce the number of transmit antennas required to achieve a given spectral efficiency, we use the structure of the transmitted vector to increase the number of spatial symbols. Let $\boldsymbol{i} = (\boldsymbol{i}_1, \boldsymbol{i}_2)$ be a spatial symbol, where the first signal symbol s_{k_1} is transmitted from the $\boldsymbol{i}_1$ combination of antennas and the second signal symbol $s_{k_2'}$ from the combination $\boldsymbol{i}_2$. For a given channel realization, the received noiseless vector $\mathbf{g}$ is given by

$$\begin{aligned} \mathbf{g} &= \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} \mathbf{h}_i + \frac{s_{k_2'}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} \mathbf{h}_i \\ &= \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} \mathbf{h}_i + \frac{s_{k_2}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} e^{j\theta} \mathbf{h}_i, \end{aligned} \tag{12}$$

where $s_{k_2'} = s_{k_2} e^{j\theta}$ and $s_{k_1}, s_{k_2} \in \Omega_a$. To obtain $\boldsymbol{i}_1$ and $\boldsymbol{i}_2$, we impose the following conditions:

(i) $\boldsymbol{i}_1 \cup \boldsymbol{i}_2 = \emptyset$: this implies that the set of antennas from which the first signal symbol and that from which the second symbol are transmitted do not overlap. This condition reduces the detection ambiguity at the receiver.

(ii) $(\boldsymbol{i}_1, \boldsymbol{i}_2) \neq (\boldsymbol{i}_2, \boldsymbol{i}_1)$: as described earlier, the first and second symbols are drawn from different constellation sets. Therefore, for given signal symbols' indices k_1 and k_2, the following two received lattice points—noiseless received vectors— are distinguishable:

$$\mathbf{g}_1 = \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} \mathbf{h}_i + \frac{s_{k_2}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} e^{j\theta} \mathbf{h}_i, \tag{13}$$

and

$$\mathbf{g}_2 = \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} \mathbf{h}_i + \frac{s_{k_2}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} e^{j\theta} \mathbf{h}_i. \tag{14}$$

Input: $\mathrm{A} = \{1, \ldots, n_T\}, n_U, n_T$
Output: The set of spatial symbols $\mathscr{I}$
$\mathscr{C} = \mathrm{COMB}(\mathrm{A}, n_U)$;
for $\boldsymbol{i}_1 \in \mathscr{C}$ **do**
  $\mathrm{B} = \mathrm{A}_{\overline{\boldsymbol{i}_1}}$;
  $\mathscr{D} = \mathrm{COMB}(\mathrm{B}, n_U)$;
  for $\boldsymbol{i}_2 \in \mathscr{D}$ **do**
    $\mathrm{APPND}(\mathscr{I}, (\boldsymbol{i}_1, \boldsymbol{i}_2))$
  end
end

Algorithm 1: Generation of the spatial symbols, assuming n_T transmit antennas and that each signal symbol is transmitted from a combination of n_U antennas.

Table 3: An example of the spatial symbols $(\boldsymbol{i}_1, \boldsymbol{i}_2)$ that can be used for transmission by GCQSM-PC, assuming $n_T = 5$ and $n_U = 2$.

$\boldsymbol{i}_1$	$\boldsymbol{i}_2$
{1, 2}	{3, 4}, {3, 5}, {4, 5}
{1, 3}	{2, 4}, {2, 5}, {4, 5}
{1, 4}	{2, 3}, {2, 5}, {3, 5}
{1, 5}	{2, 3}, {2, 4}, {3, 4}
{2, 3}	{1, 4}, {1, 5}, {4, 5}
{2, 4}	{1, 3}, {1, 5}, {3, 5}
{2, 5}	{1, 3}, {1, 4}, {3, 4}
{3, 4}	{1, 2}, {1, 5}, {2, 5}
{3, 5}	{1, 2}, {1, 4}, {2, 4}
{4, 5}	{1, 2}, {1, 3}, {2, 3}

This is rendered possible using the rotation angle θ. The optimization of the rotation angle is addressed in a following section.

Algorithm 1 depicts a pseudocode of generating the set of spatial symbols $\mathscr{I}$ that can be used for transmission. Therein $\mathrm{COMB}(\mathrm{A}, n_U)$ generates all combinations of length n_U of the elements of the set A. Also, $\mathrm{APPND}(\mathscr{I}, (\boldsymbol{i}_1, \boldsymbol{i}_2))$ appends the tuple $(\boldsymbol{i}_1, \boldsymbol{i}_2)$ to the set $\mathscr{I}$. $\mathrm{B} = \mathrm{A}_{\overline{\boldsymbol{i}_1}}$ is the difference of the two sets A and $\boldsymbol{i}_1$. Accordingly, the number of spatial symbols that can be used for transmission is given by

$$C_{\mathrm{spa}} = \binom{n_T}{n_U} \times \binom{n_T - n_U}{n_U}. \tag{15}$$

Table 3 depicts an example of the obtained spatial symbols, assuming $n_T = 5$ and $n_U = 2$. In this case, GCQSM-PC obtains 30 spatial symbols, whereas GCQSM-UC obtains only five.

Table 4 lists the number of transmit antennas required by each system to achieve a given spectral efficiency per spatial symbol. For instance, SM, GSM (and MA-SM), CQSM, GCQSM-UC, and GCQSM-PC require 1024, 46, 32, 15, and 10 transmit antennas, respectively, to achieve 10 bits per spatial symbol.

4. Performance Analysis of the Generalized CQSM

Let

$$\mathbf{g}_i = \frac{s_{k_1}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_1} \mathbf{h}_i + \frac{s_{k_2}}{\sqrt{n_U}} \sum_{i \in \boldsymbol{i}_2} e^{j\theta} \mathbf{h}_i \tag{16}$$

and

$$\mathbf{g}_k = \frac{s_{\widetilde{k}_1}}{\sqrt{n_U}} \sum_{i \in \widetilde{\boldsymbol{i}}_1} \mathbf{h}_i + \frac{s_{\widetilde{k}_2}}{\sqrt{n_U}} \sum_{i \in \widetilde{\boldsymbol{i}}_2} e^{j\theta} \mathbf{h}_i \tag{17}$$

be two noiseless received codewords corresponding to the transmitted symbol $(s_{k_1}, s_{k_2}, \boldsymbol{i}_1, \boldsymbol{i}_2)$ and $(s_{\widetilde{k}_1}, s_{\widetilde{k}_2}, \widetilde{\boldsymbol{i}}_1, \widetilde{\boldsymbol{i}}_2)$, where s_k is a signal symbol and $\boldsymbol{i} = \{l_1, \ldots, l_{n_U}\}$ is a spatial symbol. The corresponding received vectors are denoted by $\mathbf{y}_i$ and $\mathbf{y}_k$, respectively.

The conditional pairwise error probability (PEP) of the maximum-likelihood (ML) receiver is given by [21, 22]

$$\begin{aligned} \Pr\left[\mathbf{g}_i \longrightarrow \mathbf{g}_k \mid \mathbf{H}\right] &= Q\left(\sqrt{\frac{\left\|\mathbf{y}_i - \mathbf{y}_k\right\|^2}{2\sigma_n^2}}\right), \\ &= Q\left(\sqrt{\frac{\left\|\mathbf{H}\left(\mathbf{s}_i - \mathbf{s}_k\right)\right\|^2}{2\sigma_n^2}}\right), \end{aligned} \tag{18}$$

where $Q(\cdot)$ is the Gaussian tail probability function, or simply the Q-function, and $d_{i,k}^2 = \|\mathbf{y}_i - \mathbf{y}_k\|^2$ is the squared Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_k$ at the receiver. Similarly, $d_{i,k(\mathrm{tx})}^2 = \|\mathbf{s}_i - \mathbf{s}_k\|^2$ is the squared Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_k$ at the transmitter. Note that,

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}\left\{d_{i,j}^2\right\} &= \mathbb{E}_{\mathbf{H}}\left\{\left(\mathbf{s}_i - \mathbf{s}_k\right)^H \mathbf{H}^H \mathbf{H}\left(\mathbf{s}_i - \mathbf{s}_k\right)\right\} \\ &= \left(\mathbf{s}_i - \mathbf{s}_k\right)^H \left(\mathbf{s}_i - \mathbf{s}_k\right) = d_{i,k(\mathrm{tx})}^2. \end{aligned} \tag{19}$$

The unconditional PEP (UPEP), assuming n_R receives antennas, which is obtained by taking the expectation of (18) over the channel $\mathbf{H}$ is given by

$$\Pr\left[\mathbf{g}_i \longrightarrow \mathbf{g}_k\right] = \mu_{i,j}^{n_R} \sum_{l=0}^{n_R - 1} \binom{n_R - 1 + l}{l} \left[1 - \mu_{i,j}\right]^l, \tag{20}$$

where

$$\mu_{i,j} = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma \cdot d_{i,k(\mathrm{tx})}^2}{4 + \gamma \cdot d_{i,k(\mathrm{tx})}^2}}\right), \tag{21}$$

and $\gamma = \sigma_s^2/\sigma_n^2$ is the signal-to-noise ratio (SNR) with $\sigma_s^2 = \mathbb{E}(s^* s)$, the expected power of the signal symbol.

The union bound on the pairwise error probability is then given by averaging all the pairwise probabilities as follows;

$$\Pr\left[e\right] \leq \frac{1}{2^M} \sum_{i=1}^{2^M} \sum_{k=1}^{2^M} \Pr\left[\mathbf{g}_i \longrightarrow \mathbf{g}_k\right]. \tag{22}$$

Table 4: Examples of the number of transmit antennas required to achieve a given SE_{spa} for several spatial modulation systems.

SE_{spa}	SM	GSM	CQSM, QSM	GCQSM-UC	GCQSM-PC
6	64	12	8	8	6
8	256	24	16	11	8
10	1024	46	32	15	10
12	4096	92	64	20	13

Accordingly, the average bit-error rate (BER) is upper-bounded by

$$P_e \leq \frac{1}{M2^M} \sum_{i=1}^{2^M} \sum_{k=1}^{2^M} D_{\mathbf{s}_i,\mathbf{s}_k} \Pr\left[\mathbf{g}_i \longrightarrow \mathbf{g}_k\right] \tag{23}$$

where $D_{\mathbf{s}_i,\mathbf{s}_k}$, the hamming distance between $\mathbf{s}_i$ and $\mathbf{s}_k$, is the number of errors associated with the event $[\mathbf{g}_i \longrightarrow \mathbf{g}_k]$ and M is the spectral efficiency of the system. Note that in each transmitted vector symbol $\mathbf{s}$, there are exactly $(2\times n_U)$ nonzero elements.

5. Optimization of the Rotation Angles of CQSM/ICQSM Revisited

In [17, 18], the optimal rotation angles applied to the second modulation set of the CQSM and ICQSM, respectively, are obtained through extensive Monte Carlo simulations. In this case, the obtained rotation angle corresponds to the minimum bit-error rate (BER) averaged over a large number of channel and noise realizations. Alternatively, the rotation angle can be obtained through minimizing the upper-bound of the unconditional pairwise error probability.

Let

$$\begin{aligned} \mathbf{g}_i &= \mathbf{H}\mathbf{s}_i = s_{k_1}\mathbf{h}_{i_1} + s'_{k_2}\mathbf{h}_{i_2} \\ \mathbf{g}_k &= \mathbf{H}\mathbf{s}_k = s_{\tilde{k}_1}\mathbf{h}_{\tilde{i}_1} + s'_{\tilde{k}_2}\mathbf{h}_{\tilde{i}_2}, \end{aligned} \tag{24}$$

where $\mathbf{y}_i = \mathbf{g}_i + \mathbf{n}$ and $\mathbf{y}_k = \mathbf{g}_k + \mathbf{n}$ are the corresponding received vectors for a given noise vector $\mathbf{n}$.

The PEP and UPEP of the CQSM and ICQSM can also be obtained using (18)–(21). The union bound on the pairwise error probability at high SNR values (a.k.a. asymptotic probability of error) is then given by

$$\begin{aligned} \Pr[e] &\leq \frac{1}{2^M} \sum_{i=1}^{2^M} \sum_{k=1}^{2^M} \Pr\left[\mathbf{g}_i \longrightarrow \mathbf{g}_k\right] \\ &= \binom{2n_R - 1}{n_R} \frac{\gamma^{-n_R}}{2^M} \sum_{i=1}^{2^M} \sum_{k=1}^{2^M} \left(d^2_{i,k(\mathrm{tx})}\right)^{-n_R} \\ &= \binom{2n_R - 1}{n_R} \frac{\gamma^{-n_R}}{L^2} \sum_{i=1}^{B} f_i \Omega_i, \end{aligned} \tag{25}$$

where the maximum value of B equals 2^6 as shown in [19]. The term Ω_i is a function of the signal symbols $(s_{k_1}, s'_{k_2}, s_{\tilde{k}_1}, s'_{\tilde{k}_2})$ and n_R, and f_i is the frequency of Ω_i, where $\sum_{i=1}^{B} f_i = n_T^4$. It is shown in [19] that $B = 7$ in the case of ICQSM and it can be shown to be equal to 15 for CQSM. The rotation angle is then obtained by solving the following optimization problem:

$$\underset{0<\theta<\pi/2}{\operatorname{argmin}} \left(\sum_{i=1}^{B} f_i \Omega_i \right), \tag{26}$$

where $B = 15$ and 7 for CQSM and ICQSM, respectively.

Figure 1 depicts the cost function of (26) for CQSM and ICQSM and for several system configurations. The optimal rotation angle corresponds to the minimum value of the cost function. Assuming the case of CQSM and for the scenarios depicted in the figure, the obtained optimal rotation angles are *very close* to those obtained in [17]. Therefore, this small and tolerable degradation in the error performance comes at a high gain in the optimization time of the rotation angle. On the other hand, the obtained rotation angles for ICQSM are identical to those obtained in [18]. Assuming PSK modulation, the optimal rotation angle for the ICQSM is equal to π/L, with L denoting the cardinality of the constellation set.

6. Optimization of the Rotation Angles for the Generalized CQSM

6.1. Rotation Angle of the GCQSM-UC. As explained earlier and based on the design of the spatial symbols of the GCQSM-UC, if $(\boldsymbol{i}_1, \boldsymbol{i}_2)$ is a spatial symbol that can be used for transmission then $(\boldsymbol{i}_2, \boldsymbol{i}_1)$ is not a valid spatial symbol. This implies that the first symbol s_{k_1} and s'_{k_2} are distinguishable using the indices of the antennas from which they are sent. As such, both symbols can be drawn from the same modulation set and hence the rotation angle will have no impact on the performance of the system. In the following θ is set to zero for GCQSM-UC.

6.2. Rotation Angle of the GCQSM-PC. Opposed to the GCQSM-UC, both $(\boldsymbol{i}_1, \boldsymbol{i}_2)$ and $(\boldsymbol{i}_2, \boldsymbol{i}_1)$ are valid spatial symbols that can be used for transmission in GCQSM-PC. Therefore, the second symbol s'_{k_2} should be rotated so that it can be distinguished from the first signal symbol s_{k_1}. The formulation of the upper-bound of the GCQSM-PC as in (22) is a hard problem. Alternatively, we obtained the rotation angle using simulations. Figure 2 depicts the BER of GCQSM-PC versus the rotation angle: the upper subfigure assumes QPSK modulation and the lower subfigure assumes 16-QAM. The optimal rotation angle is about $\pi/4$ for all the simulated scenarios assuming QPSK modulation. Also, the optimal

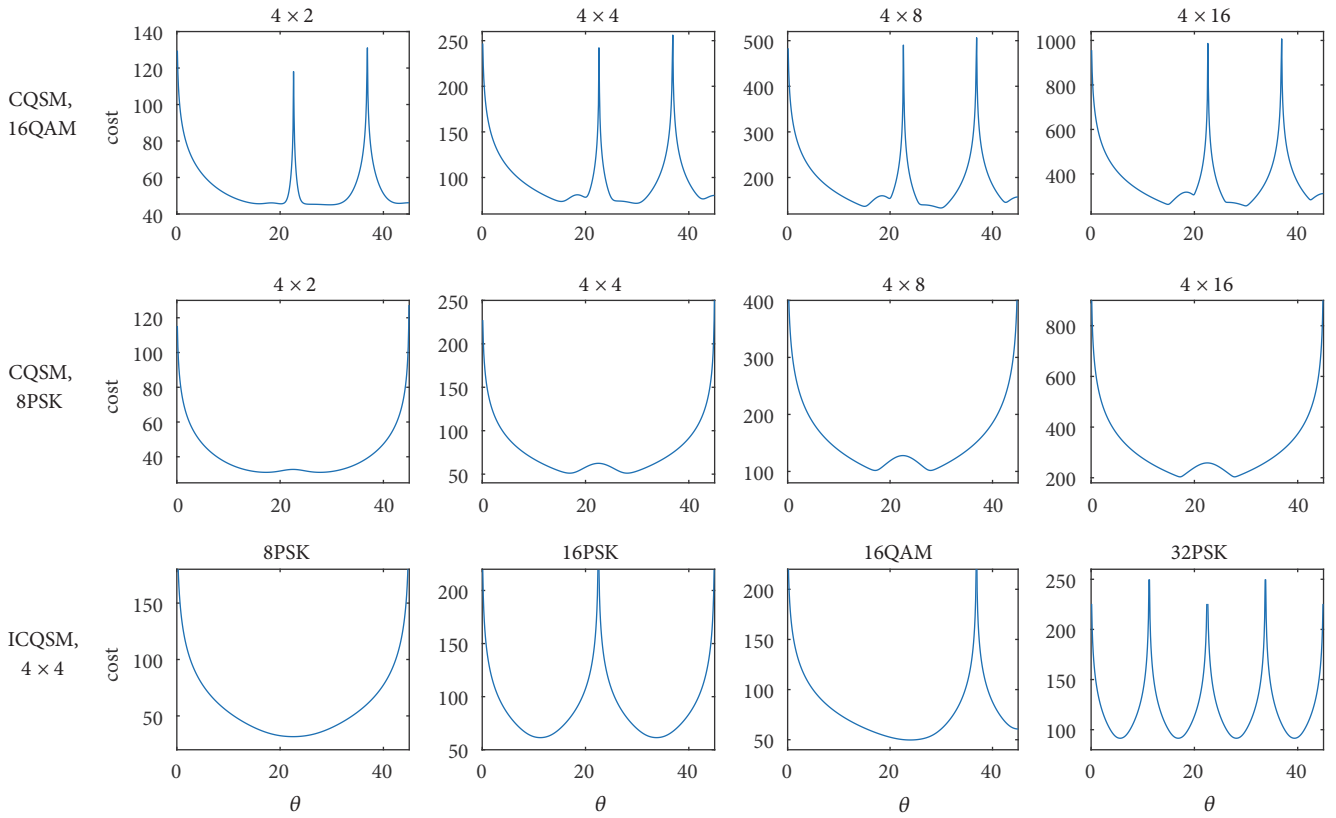


Figure 1: The cost function in (26) in dB scale versus the rotation angle for several system configurations. The subfigures in the first and second row are obtained for CQSM and those in the bottom row are for ICQSM.

rotation angle for the simulated scenarios using 16-QAM is approximately 23 degrees. These values of the optimal rotation angle are affected by the sets from which symbols are drawn, and the values of n_T and n_R. As a future work, we would like to analytically derive a generic formula to obtain the optimal angle for arbitrary system configurations.

7. Simulation Results

The following results are obtained assuming that the channel state information (CSI) is known only at the receiver. To make a fair comparison, the rotation angles for the simulated scenarios are optimized using simulations for both CQSM and the generalized CQSM. The obtained rotation angles using simulations and the optimization method described in Section 5 for CQSM scheme have very similar values.

Figure 3 depicts comparisons between the MA-SM and the CQSM for several system configurations of (n_T, q) and a fixed $n_R = 4$. We make the following comparisons.

(i) *For the same n_T and q and different spectral efficiency (Figure 3(a))*: in this case, MA-SM outperforms CQSM by about 2 dB for $n_T = 8$ and about 1.5 dB for $n_T = 16$. The gap is bridged between the two algorithms as n_T becomes large. This decrease in the performance gap is due to the decrease in the probability of transmitting the two symbols from the same transmit antenna in the CQSM. This probability is equal to $1/n_T$. The comparison here is not fair because the CQSM achieves a spectral efficiency 2 bpcu higher than that of the MA-SM.

(ii) *For the spectral efficiency and q and different n_T (Figure 3(b))*: the performance of the two algorithms is depicted for an equal spectral efficiency assuming $n_T = 8, 16$ and $n_T = 12, 24$ for CQSM and MA-SM, respectively. For these two scenarios, CQSM requires 4 and 8 transmit antennas less than MA-SM. This huge reduction in the number of transmit antennas comes at a negligible degradation in the SNR for high n_T. This implies that the CQSM is more attractive for massive MIMO systems.

(iii) *For the spectral efficiency and n_T and different q (Figure 3(c))*: finally, the performance of the two algorithms is evaluated for the same spectral efficiency and number of transmit antennas. In this case, q is set to 2 and 3 in the CQSM and MA-SM, respectively. At a target BER of 10^{-4}, CQSM outperforms MA-SM by 2 and 2.4 dB, respectively, for the two depicted system configurations.

According to this analysis and results depicted in Figure 3, we conclude that CQSM can be used to reduce the number of transmit antennas at a marginal cost in the SNR for high n_T or it can achieve an improvement in the SNR if the same number of antennas is deployed by the two systems.

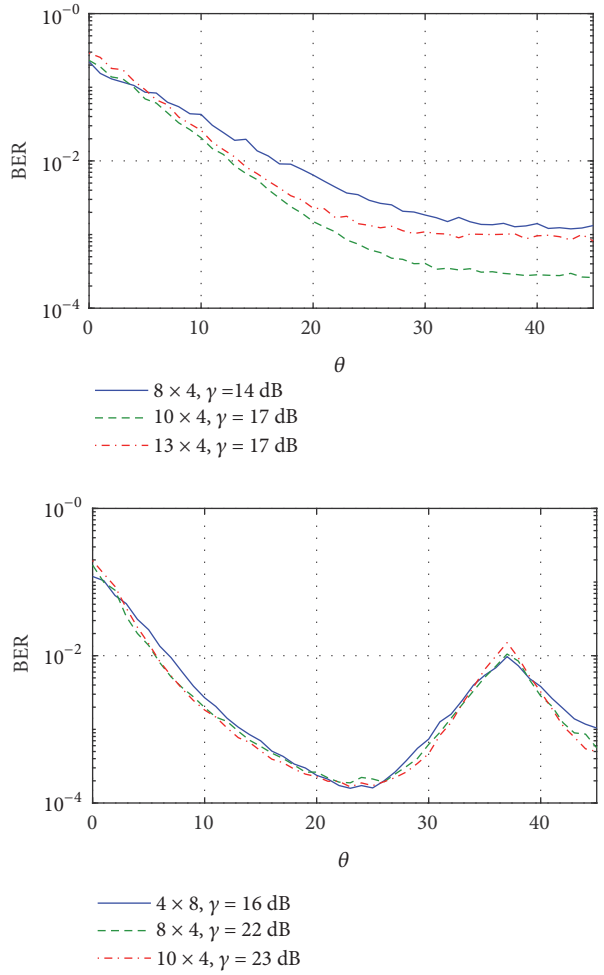


FIGURE 2: BER of the GCQSM-PU versus the rotation angle applied to the second signal symbol, using QPSK (first subfigure) and 16-QAM (second subfigure).

Since at high n_T CQSM requires one antenna less than ICQSM to achieve the same spectral efficiency at the cost of a moderate degradation in the BER performance, CQSM is used in the following comparisons.

Figure 4 depicts the BER performance of CQSM, GCQSM-UC, and GCQSM-PC for several system scenarios and $n_U = 2$. The title of each subfigure is a tuple that includes the number of transmit antennas required by CQSM, GCQSM-UC, and GCQSM-PC, respectively, to achieve the same spectral efficiency. The spectral efficiency SE_{spa} achieved assuming the system's parameters in Figure 4(a) is 6, in 4(b) is 8, in 4(c) is 10, and in 4(d) is 12 bpcu. For each of the system configurations SE_{sig} is equal to 4 bpcu, leading to a total spectral efficiency $\mathrm{SE} = \mathrm{SE}_{spa} + \mathrm{SE}_{sig}$ of 10, 12, 14, and 16 bpcu, respectively. The range over which the rotation angle is optimized is $[0, \pi/4]$. The rotation angles applied to the second signal symbol for the scenarios depicted in Figure 4 are listed in Table 5. The rotation angles are optimized through Monte Carlo simulations. Based on the results depicted in Figure 4, we make the following remarks in terms of the BER performance, the rotation angles, and the number of employed transmit antennas.

(i) For $n_R = 2$, the three techniques achieve the same BER performance regardless of the number of transmit antennas.

(ii) For a given value of n_R, the BER performance is degraded as the number of transmit antennas increases. The degradation is relatively small as n_R increases. This is an expected behavior of SM techniques because the size of hypothesis set over which the ML detector performs the search increases as n_T increases.

(iii) In Figure 4(a), GCQSM-PC slightly outperforms the other two techniques for $n_R = 4$ and 8. As n_T gets larger, the probability that the two signal symbols are transmitted from the same antenna in CQSM reduces. Accordingly, the performance of CQSM and the two generalized algorithms almost coincide for $n_R = 4$ in Figures 4(b), 4(c), and 4(d). For $n_R = 8$, CQSM slightly outperforms GCQSM-UC and GCQSM-PU.

(iv) As shown in Figure 4(c), the generalized techniques perform close to the CQSM for $n_R = 4$. In this case, GCQSM-UC and GCQSM-PC require 17 and 22

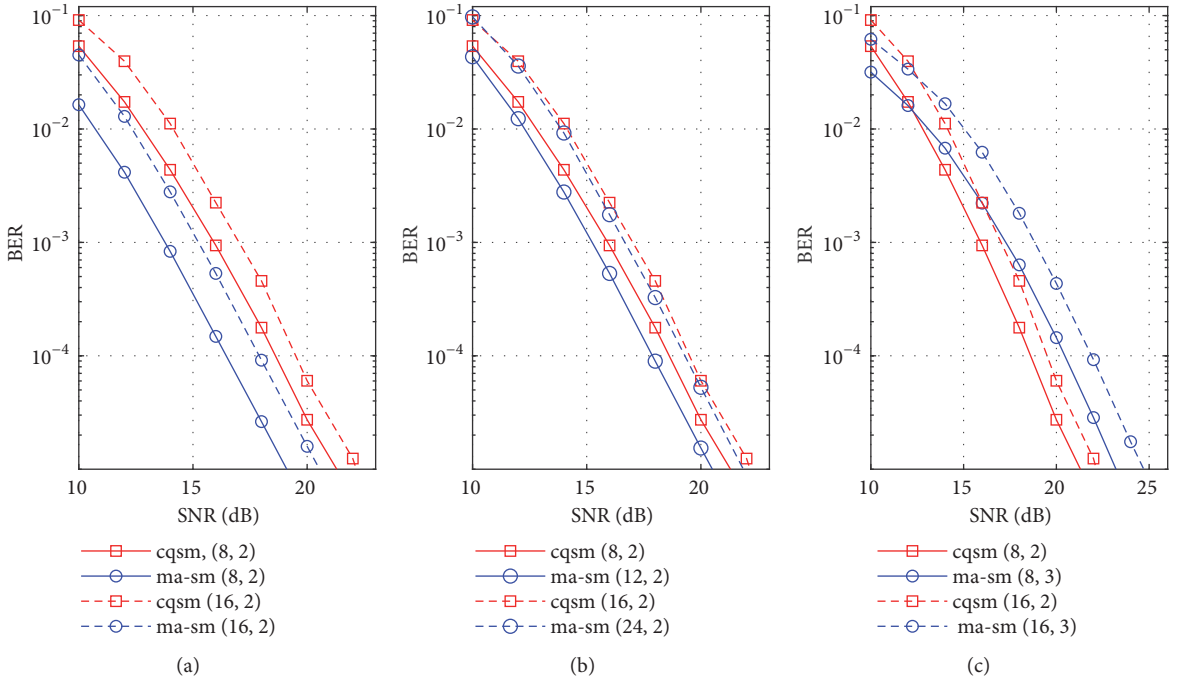


FIGURE 3: A comparison between CQSM and MA-SM: (a) for the same n_T and different spectral efficiency, (b) for the same spectral efficiency and q, and different n_T, and finally (c) for the same spectral efficiency and n_T, and different q.

TABLE 5: Range of optimal rotation angles used to obtain the BER results depicted in Figure 4, assuming QPSK modulation.

SE_{spa}	n_R	θ_{CQSM}	$\theta_{GCQSM-PC}$
6	2	≥25	≥30
	4	≥35	≥41
	8	≥38	≥43
8	2	≥20	≥21
	4	≥34	≥35
	8	≥38	≥42
10	2	≥9	≥12
	4	≥30	≥32
	8	≥38	≥39
12	2	≥9	≥12
	4	≥25	≥30
	8	≥38	≥39

transmit antennas less than CQSM while achieving the same spectral efficiency.

Finally, we compare the BER performance of SM, GSM, QSM, CQSM, and the proposed generalized CQSM techniques. We assume that all the schemes achieve the same ES_{spa} and ES_{sig}, leading to the same total spectral efficiency of $ES = (ES_{spa} + ES_{sig})$. Also, the three generalized schemes use a combination of two antennas to transmit each signal symbol. Figures 5(a) and 5(b) depict the BER performance for a spectral efficiency of 16 and 18 bpcu, respectively. The number of transmit antennas and bits per signal symbol is represented as a tuple of the form (n_T, q) for each of the systems. The rotation angles applied to CQSM and GCQSM-PC are 20 and 23 degrees, respectively, to obtain the results in the two subfigures. We make the following two remarks:

(i) CQSM, GCQSM-PC, and GCQSM-UC apply 16-QAM modulation, whereas SM, GSM, and QSM use 256-QAM to achieve an ES_{sig} of 8 bpcu. The number of transmit antennas required to achieve the spectral efficiency of 16 and 18 bpcu is given in the second and third row, respectively, in Table 4. GCQSM-PC requires the least number of transmit antennas of 8 and 10 for the results in Figures 5(a) and 5(b), respectively. On the other hand, SM requires 256 and

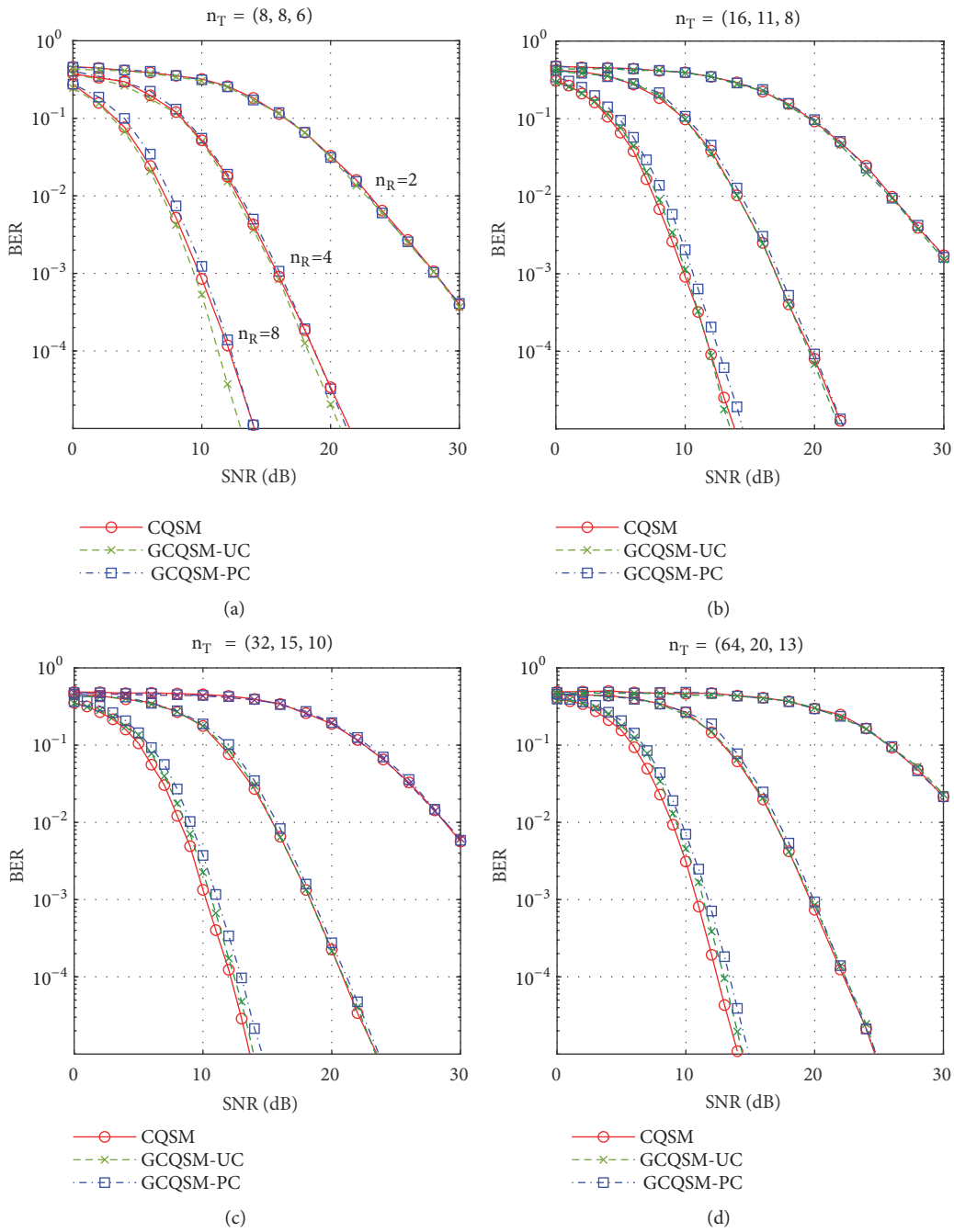


Figure 4: BER performance of CQSM, GCQSM-UC, and GCQSM-PC using QPSK modulation, where the spectral efficiency is equal to (a) 10, (b) 12, (c) 14, and (d) 16 bpcu. The title of each subfigure is a tuple of the number of antennas required by CQSM, GCQSM-UC, and GCQSM-PC, respectively.

1024 antennas, respectively, to achieve the same SE. Compared to CQSM, GCQSM-PC requires 8 and 22 less antennas to achieve the same SE_{spa} of 8 and 10 bpcu, respectively.

(ii) For both scenarios, GCQSM-UC achieves the best performance followed by GCQSM-PC and CQSM. In Figure 5(b), GCQSM-UC outperforms GCQSM-PC and CQSM by 1 and 1.6 dB, respectively. However, GCQSM-PC requires 5 less transmit antennas than GCQSM-UC. Therefore, this slight degradation in the SNR is tolerable in the case of GCQSM-PC given the high reduction in the number of transmit antennas used to achieve the same spectral efficiency.

Future Work. We would like to investigate the case of generalized CQSM where the number of transmit antennas used for transmitting each of the two signal symbols can be

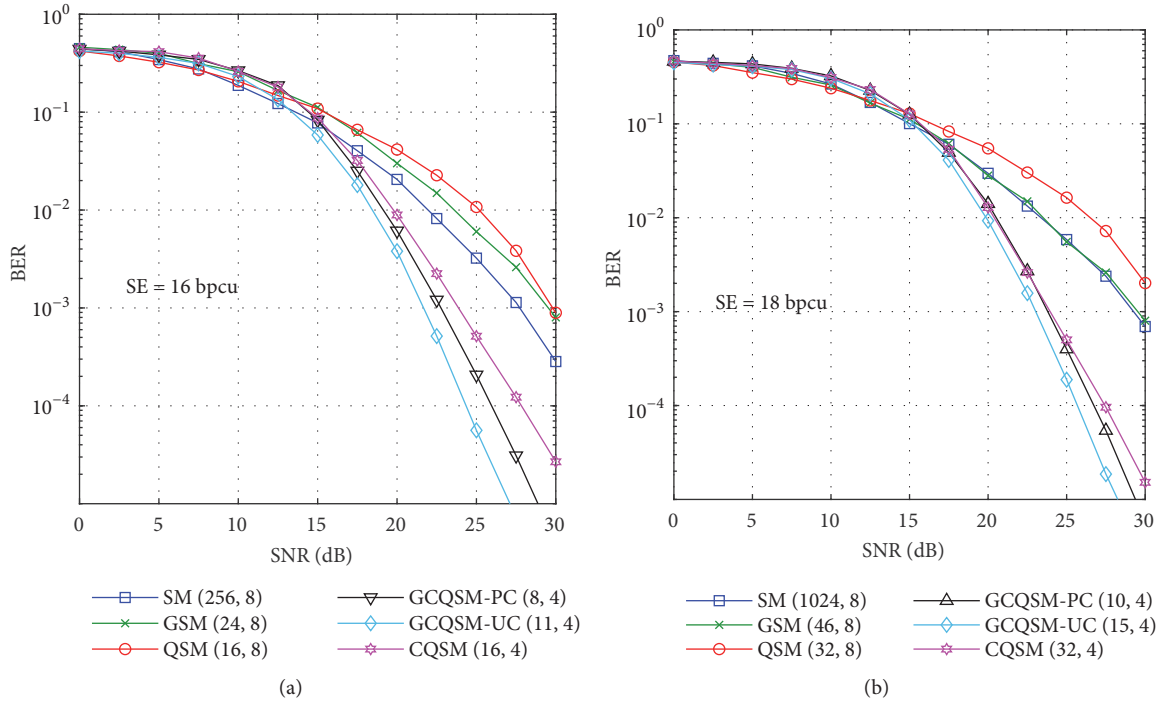


Figure 5: BER comparison between the generalized CQSM schemes, CQSM, SM, GSM, and QSM for spectral efficiency of (a) 16 and (b) 18 bpcu. CQSM, GCQSM-UC, and GCQSM-PC use 16-QAM and the remaining schemes use 256-QAM.

variable. Accordingly, the total number of spatial symbols will be increased, leading to higher spectral efficiency.

8. Conclusions

In this paper, we introduced two generalizations of the complex quadrature spatial modulation, namely, GCQSM with unique combinations (GCQSM-UC) and with permuted combinations (GCQSM-PC). While the former generalization uses the conventional spatial symbol generation of the GSM scheme, the later scheme relies on the fact that the second signal symbol is distinguishable from the first through the angle rotation. This allows expanding the set of antenna combinations that can be used for transmission, leading to the reduction in the number of transmit antennas required to achieve a given spectral efficiency. The two proposed generalizations perform close to CQSM using QPSK and outperform it using higher order modulation schemes. Also, GCQSM-PC requires 10 antennas to achieve a spectral efficiency of 10 bcpu per spatial symbol. To achieve the same efficiency, GSM requires 46, and CQSM and QSM require 32 transmit antennas.

Data Availability

No data were used to support this study.

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 stipend by Korea Tech in the period 2018-2019.

References

[1] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693–16746, 2017.

[2] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 years of permutation, spatial and index modulation: from classic RF to visible light communications and data storage," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1905–1938, 2018.

[3] R. Y. Mesleh, H. Haas, S. Sinanović, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[4] G. Kaddoum, M. F. A. Ahmed, and Y. Nijsure, "Code index modulation: a high data rate and energy efficient communication system," *IEEE Communications Letters*, vol. 19, no. 2, pp. 175–178, 2015.

[5] J. Zhang, Y. Wang, J. Zhang, and L. Ding, "Polarization shift keying (PolarSK): system scheme and performance analysis," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10139–10155, 2017.

[6] E. g. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, 2013.

[7] X. Zhu, Z. Wang, Q. Wang, and H. Haas, "Virtual spatial modulation for mimo systems," in *Proceedings of the IEEE*

Global Communications Conference (GLOBECOM '16), pp. 1–6, Wash, DC, USA, 2016.

[8] J. Li, M. Wen, M. Zhang, and X. Cheng, "Virtual spatial modulation," *IEEE Access*, vol. 4, pp. 6929–6938, 2016.

[9] P. Ju, M. Zhang, X. Cheng, C. Wang, and L. Yang, "Generalized spatial modulation with transmit antenna grouping for correlated channels," in *Proceedings of the IEEE International Conference on Communications (ICC '16)*, pp. 1–6, Kuala Lumpur, Malaysia, 2016.

[10] S. AbuTayeh, M. Alsalahat, I. Kaddumi, Y. Alqannas, S. Althunibat, and R. Mesleh, "A half-full transmit-diversity spatial modulation scheme," in *Proceedings of the Broadband Communications, Networks, and Systems. 9th International EAI Conference, Broadnets '18*, Faro, Portugal, 2018.

[11] A. Younis, R. Mesleh, M. Di Renzo et al., "Generalized spatial modulation for large-scale MIMO," in *Proceedings of the IEEE Communications Society*, pp. 346–350, Eusipco, Lisbon, Portugal, 2014.

[12] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation with multiple active transmit antennas and low complexity detection scheme," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1605–1615, 2012.

[13] T. L. Narasimhan, P. Raviteja, and A. Chockalingam, "Generalized Spatial Modulation in Large-Scale Multiuser MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3764–3779, 2015.

[14] R. Mesleh, S. S. Ikki, and H. M. Aggoune, "Quadrature spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2738–2742, 2015.

[15] F. R. Castillo-Soria, J. Cortez-González, R. Ramirez-Gutierrez, F. M. Maciel-Barboza, and L. Soriano-Equigua, "Generalized quadrature spatial modulation scheme using antenna grouping," *ETRI Journal*, vol. 39, no. 5, pp. 707–717, 2017.

[16] R. Mesleh, O. Hiari, and A. Younis, "Generalized space modulation techniques: hardware design and considerations," *Physical Communication*, vol. 26, pp. 87–95, 2018.

[17] M. Mohaisen and S. Lee, "Complex quadrature spatial modulation," *ETRI Journal*, vol. 39, no. 4, pp. 514–524, 2017.

[18] M. Mohaisen, "Increasing the minimum Euclidean distance of the complex quadrature spatial modulation," *IET Communications*, vol. 12, no. 7, pp. 854–860, 2018.

[19] A. Iqbal, M. Mohaisen, and K. S. Kwak, "Modulation set optimization for the improved complex quadrature SM," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 6769484, 12 pages, 2018.

[20] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, vol. 45, Springer, Berlin, Germany, 2008.

[21] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2nd edition, 2004.

[22] M. Di Renzo and H. Haas, "Bit error probability of SM-MIMO over generalized fading channels," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 3, pp. 1124–1144, 2012.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 5287679, 8 pages
https://doi.org/10.1155/2019/5287679

Research Article

A Compact Ku-Band Active Electronically Steerable Antenna with Low-Cost 3D T/R Module

Zheng Xu,[1] Chengxiang Hao,[1] Kuiwen Xu,[2] and Shichang Chen [2]

[1] *The University of Chinese Academy of Sciences, Beijing 101400, China*
[2] *The Key Lab of RF Circuits and Systems of the Ministry of Education of China, Hangzhou Dianzi University, Hangzhou 310018, China*

Correspondence should be addressed to Shichang Chen; eechensc@hdu.edu.cn

Received 9 November 2018; Revised 21 March 2019; Accepted 14 April 2019; Published 22 April 2019

Guest Editor: Sungchang Lee

This paper presents a novel compact Ku-band active electronically steerable antenna array design with a low-cost and integrated T/R 3D module employed for airborne synthetic aperture radar (SAR) systems. The entire system adopts 3D multilayer technology with vertical interconnection to construct the hermetically packaging RF modules. By assembling different multifunctional modules into a whole multilayer board, the 3D T/R technique greatly improves the system integration and reduces implementation cost and size. Besides, a wideband circular polarized antenna array was designed in LTCC and connected to the proposed T/R modules to form a complete AESA. The whole proposed antenna system has been fabricated and experimentally investigated. Measurement results showed very good phased array performances in terms of gain, axial ratio, and radiating patterns. The low-cost, lightweight, and low-power features exhibited by the proposed design validate its applicability for weight and power constrained platforms with great electronic steering ability.

1. Introduction

Radar imaging is widely used in applications such as geographical mapping [1], security surveillance [2], and collision avoidance [3]. Among many diversified solutions, the synthetic aperture radar (SAR) is a very powerful one because it can synthesize a large aperture to yield increased resolution by using coherent processing of multiple observations from a moving antenna [4].

On the other hand, due to the quick advancements of airborne techniques represented by the unmanned aerial vehicle (UAV) recently, SAR systems start to appear in some airborne platforms. This significantly extends the mobility, flexibility, and economic efficiency of this special radar. It is well known that the successful implementation of a SAR demands a high performance active electronically steerable antenna (AESA), which is a vital and widespread technique that is commonly utilized in various military and civil applications, such as weather monitoring [5], aircraft surveillance [6], space satellite communication [7], and imagining [8]. To meet the stringent requirements of power and loading capacity for an airborne system, the AESA must be designed as compact and lightweight. As the most important components in an AESA, the transmit/receive module (TRM) and radiating antenna array are therefore demanded to be power efficient, broadband, dimensionally small, and widely steerable.

To reduce the footprint of TRM, a preliminary 3D packaging concept that vertically stacks multiple submodules has been proposed in the literature [9–11]. Some well-established processes represented by the low temperature cofired ceramics (LTCC), which are particularly suitable for the manufacture of RF modules, are often applied to assemble several MMICs or discrete chips and that can be fabricated with low cost [12, 13]. Passive components can be placed at various layers, and 3D integration enables the compactness of modules. Besides, owing to the high dielectric constant and the low dielectric and conductor losses, the LTCC technology is also widely used in the design of millimeter-wave antenna arrays [14–16]. Moreover, given the advantages of small size, low profile, and light weight, the patch antennas with LTCC technique are getting more attention for phased array implementations [16, 17].

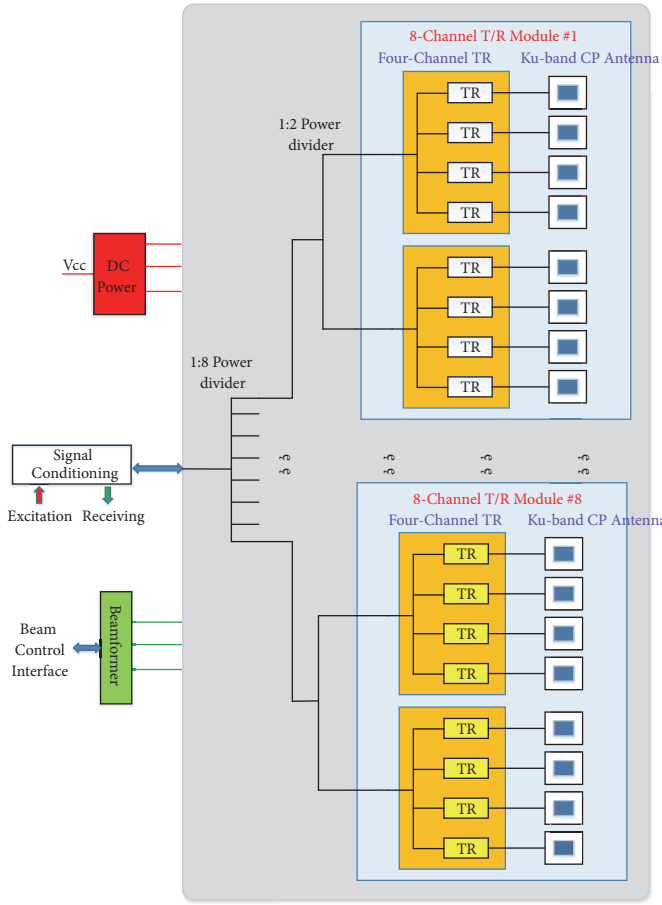


Figure 1: Schematic diagram of the proposed active electronically steerable antenna (AESA) system.

In this work, a Ku-band AESA with low cost and compact T/R 3D module for airborne SAR systems is devised with a reduced overall implementation cost and rapid prototyping speed. The remainder of this paper is organized as follows: the whole system is briefly introduced in Section 2. The design details of the TRM and antenna array are given in Section 3. Section 4 shows the simulation and measurement results of the devised system. Section 5 finally gives the conclusion.

2. System Overview

2.1. AESA Architecture. The main motivation of this work is to design a compact and cost-effective AESA system intended for airborne synthetic aperture radar in Ku-band. The main electrical and mechanical specifications are listed as follows:

(1) operation frequency: 15~16 GHz

(2) array size: 8 × 8

(3) beam width: 12°

(4) scan range: ±60° @azimuth direction, ±60°@range direction

(5) polarization: right-handed circular

(6) axial ratio: ⩽3 dB

(7) beam switching time: ⩽20 uS

(8) beam steering error: ⩽0.5°

(9) EIRP: ⩾24.4 dBW

(10) size: 100 × 100 × 50 mm3

(11) maximum power: ⩽100 W

(12) antenna weight: ⩽1 Kg.

Figure 1 shows the schematic diagram of the proposed AESA system. According to the listed requirements above, the AESA is designed at the center frequency of 15.5 GHz, with a 1 GHz bandwidth. It is mainly comprised of eight active sub-modules; each module constitutes an eight-channel 3D T/R module which drives a Ku-band circular polarized antenna array. The antenna features a rectangular grid alignment; the distance between each radiating element is 9.5 mm (roughly $\lambda_g/4$), in both azimuth and range directions.

The architecture of the whole system can be divided into three distinct layers. The top one is the radiating layer, where a compact and low-profile microwave antenna array is designed. The middle layer is intended for T/R modules and heat dissipation. The T/R module is aligned in the azimuth direction, containing eight channels. The entire AESA has eight parallel T/R modules in the range direction. This middle layer is connected to the radiating layer through the RF

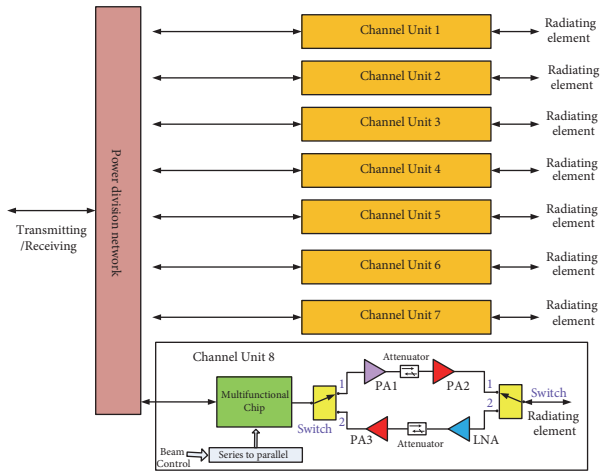


FIGURE 2: Diagram of the proposed AESA system.

SSMP connectors, with the help of structural fixture. The bottom layer is reserved for a multifunctional board which contains feeding network, beam control, power distribution, and so on. It connects to the middle layer through the RF SSMP and low-frequency connectors, providing beam control instructions, DC power, and RF signal.

3. Module Design

3.1. Low-Cost 3D T/R Module. As it is well known, the T/R module is one of the most crucial modules in a front end. It produces a high-power signal that drives antenna in the transmitting mode and amplifies weak signal got from antenna for further processing in the receiving mode. The selection of T/R module solution should comprehensively consider factors like electrical characteristics, hardware reliability, maintainability, and cost. Besides, the channel number, interface type, and structure should also be optimized in order to achieve the best overall array performances.

Figure 2 shows the diagram of the devised T/R module. In this particular work, the proposed T/R module integrates eight channels. A one-to-eight power divider is included; every output terminal drives an independent T/R unit (depicted in the zoomed-in insert), which is further connected to a radiating antenna array. Each single unit physically contains a receiving path, a transmitting path, and a common path. To achieve large system integrity and guarantee reliability, the transmitting and receiving paths share a common multifunctional chip which integrates a 6-bit phase shifter and frequency synthesizer. The link budget (signal level budget) of the front end is carefully calculated in the Keysight SystemVue software [18]. Transmission line and relevant interconnect losses are considered together with the gain characteristics of chips. Deliberate tradeoffs are made to ensure a best system performance. Meanwhile, adequate design tolerances are reserved to prevent possible failure due to component variation.

As the low-noise amplifier (LNA) and power amplifier (PA) elements have different RF performance specifications, discrete chips are used and connected with the multifunctional chip.

As known to all, the noise figure and gain of first stage largely affect the overall noise performance of the receiving channel. As the receiving channel requires very high gain, a two-stage configuration is adopted for the LNA design. A GaAs LNA chip with 1.5 dB noise figure is used for the first stage. Because the total loss of interconnect, switch, and bonding wire is about 1 dB, the overall noise figure is therefore 2.5 dB. Regarding the transmitting channel, the combination of a GaAs PA and driver chip is applied. Herein, to have good amplitude and phase balance, a 6-bit digitally controlled attenuator is added to the T/R module along with the high resolution phase shifter in the multifunctional chip. Figure 3 depicts the basis specification distribution diagram of the proposed AESA.

In order to reduce system implementation cost and save circuit size and weight as well as prototyping time, a cost-effective multilayer PCB technique is used. To be specific, a microwave board and an analog board stick together to form a single board physically. The microwave board contains all the RF components, interconnects, and transitions. RF component includes but not limited to LNA, PA, phase shifter, multifunctional chip, filters, and multiplexer. The analog board, on the other hand, integrates the low-frequency components such as power management IC and signal processing unit, to name a few. With the proposed multilayer PCB technology, it is possible that carefully designed microwave and analog boards can be assembled directly, obviating the need for much shielding between each other which is conventionally added for electromagnetic inference suppression. In the meantime, it helps us to improve installation speed and hardware reliability for mass production.

Figure 4 shows the exploded view of the assembled T/R module. The multilayer board is inserted and fixed to the cavity, and a metal plate is placed on top of it for mechanical protection. Nine SMP connectors and a 21-pin low-frequency socket are designed at the two sides of the module.

In order to improve the stability of the devised system, several treatments are applied. Firstly, a time division DC regulation mechanism is adopted. To be specific, the transmitting and receiving circuitries never operate at the same time. By doing so, mutual interference between the two modes is greatly mitigated. Secondly, comprehensive electromagnetic and simulations are conducted and metallic walls and absorptive materials are added to prevent unwanted cavity resonance. Thirdly, strip lines are widely used to isolate microwave circuits from digital and low-frequency analog circuits. The overall dimension of the fabricated eight-channel T/R module is 83 mm × 24 mm × 9.3 mm. Figure 5 gives the 3D assembly view of the whole structure.

3.2. Thermal Management. Proper thermal design is of crucial importance for space application in which unwanted heat should be dissipated efficiently. For first contribution to this issue, two fans are applied at the air inlet to have a supply air rate of 84 m^3/h and a maximum pressure of 560 Pa. Besides, a series of fins are arranged in each TR air duct to take away the heat. The thickness of fins is thinned to 0.75 mm, while

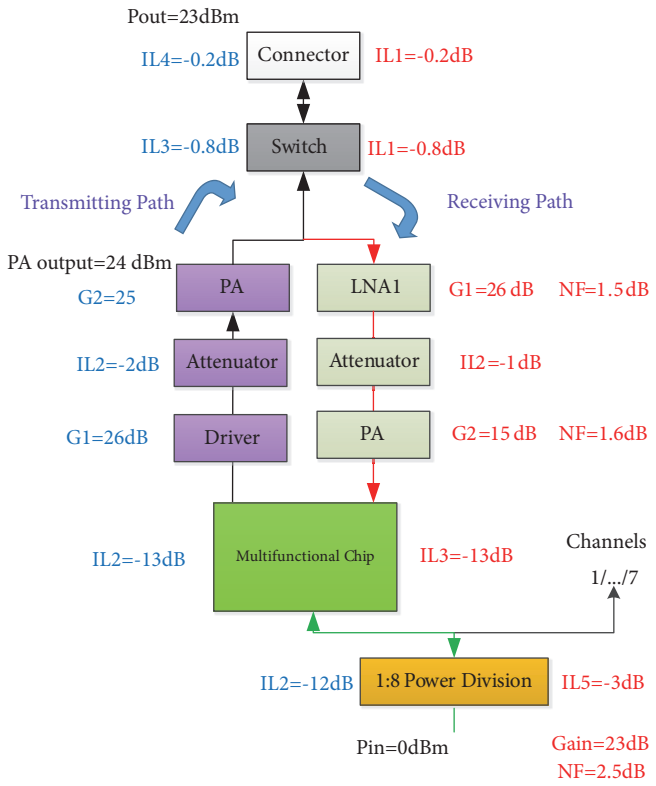


Figure 3: Basic system specification distribution diagram of the proposed AESA.

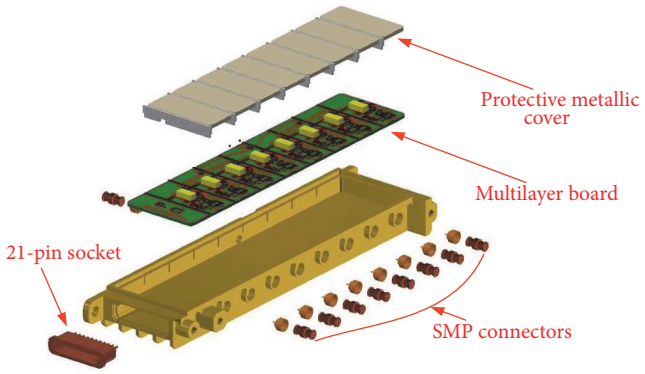


Figure 4: Exploded view of the design of 3D T/R module.

the spacing between adjacent fins is optimized to be 2 mm. These two strategies help us to improve dissipation area. Moreover, the T/R and power modules are soldered directly to the metallic structure using AuSn alloy instead of epoxy resin. For better demonstration, thermal simulations are conducted in ANSYS Workbench to validate the above design considerations. Figure 6(a) shows the simulated temperature distribution of the package. For an ambient temperature of 40°C, the peak chip temperature is around 68°C. Figure 6(b) depicts the transient thermal analysis result. As one can see, the maximum temperature can be achieved and stabilized in roughly 300 seconds after system starts up, fully validating the effectiveness of heat design.

3.3. Ku-Band Antenna Array Design on LTCC. In order to keep the entire active electronically steerable antenna compact, light, and easy to integrate and install, a patch antenna array composed of 8 × 8 radiated elements is constructed on the top of AESA with a microwave LTCC substrate. A microstrip power divider with 90° phase difference outputs as well as the slot coupling technique is used to feed each element for realizing right-handed circular polarization [19]. The proposed antenna array is designed individually and connects with the T/R module via SSMP. The circularly polarized radiator is designed to operate at 15.5 GHz with a bandwidth of more than 1 GHz, which is fed by a probe along the diagonal line of the rectangular patch. The distances along the azimuth and range directions between the elements are 9.5 mm, which is about half of wavelength at 15.5 GHz so as to avoid strong coupling. Each radiator is independently controlled by a suit of T/R module so that the amplitude and the phase of the excitation of each radiator can be tuned continuously.

4. Simulation and Measurement Results

To demonstrate the functionality of the proposed AESA, a prototyping system is manufactured and tested. Figure 7 shows a photograph of the fabricated antenna array. The devised systems are measured in a microwave anechoic chamber. A remote-controlled laptop is connected to it, which conveniently sends the required configuration commands to the systems.

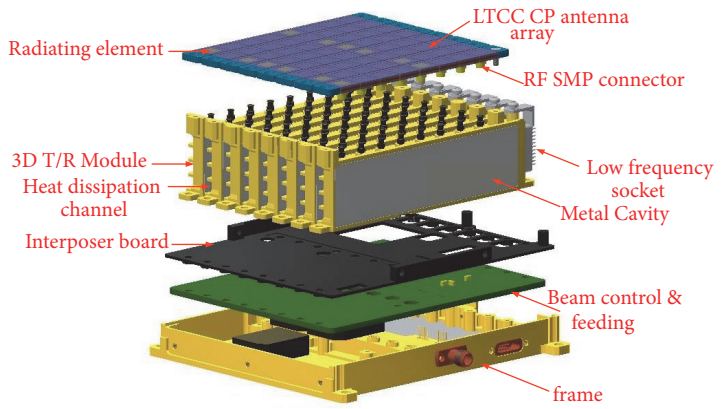
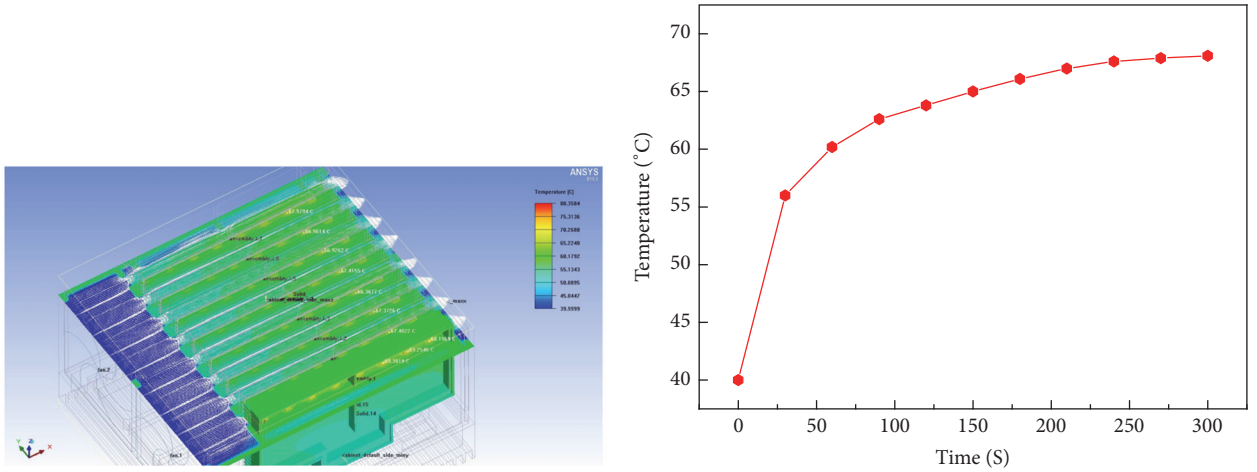


FIGURE 5: 3D assembly view of the whole structure.



(a) (b)

FIGURE 6: Thermal simulation results: (a) package temperature distribution and (b) transient thermal response.

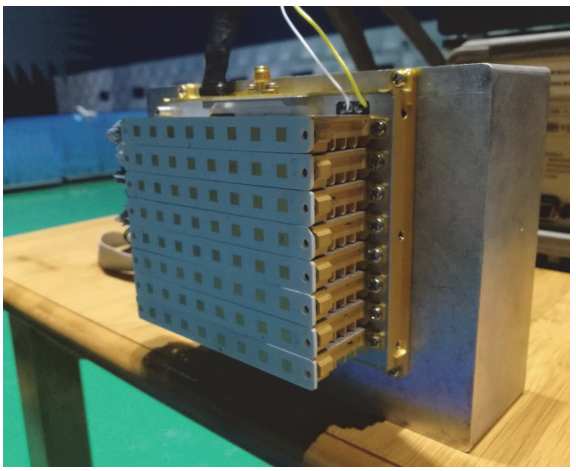


FIGURE 7: Photograph of the implemented ASEA.

The proposed 8 × 8 array antenna is simulated and optimized with HFSS. And the voltage standing wave ratios (VSWRs) of the proposed AESA (excluding the active parts) are measured using Agilent E8363B vector network analyzer.

As indicated in Figure 8, most of the simulated and measured VSWRs are below 1.5 at the operating band (from 15 GHz to 16 GHz) and good agreements are achieved between the two sets of results, where the slight differences are still acceptable and are most likely caused by the discrepancy in the permittivity of the substrate, the unbalanced coaxial cable feeding, and the uncertain factors during manufacturing. These results prove that the proposed AESA exhibits a good electrical performance.

The measured and simulated axial ratios (ARs) are depicted in Figure 9. It is readily seen that the proposed antenna achieves a 0.5 GHz bandwidth with the axial-ratio bandwidth of AR < 3 dB (from 15.25 GHz to 15.75 GHz). There is some discrepancy between the simulated and the measured results owing to some unbalanced phase shifts in the fabricated feeding network. However, the bandwidth of the simulated and measured results tends to be of the same level.

The simulated and measured radiation patterns at the center frequency, i.e., 15.5 GHz, at different steering angles,

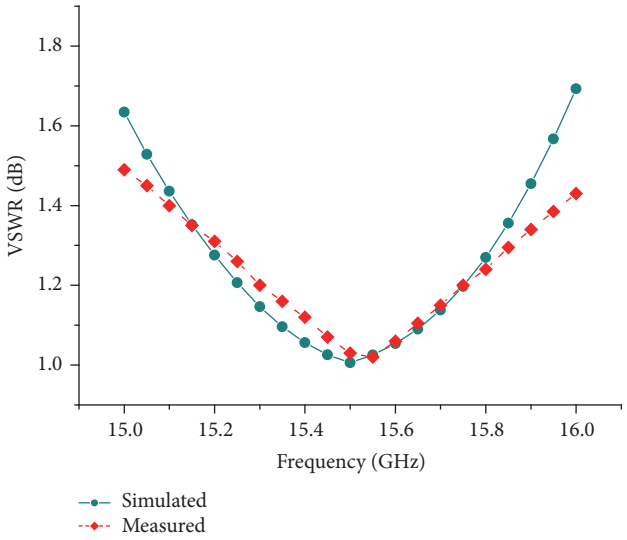


FIGURE 8: Return loss of the proposed ASEA (simulated and measured results).

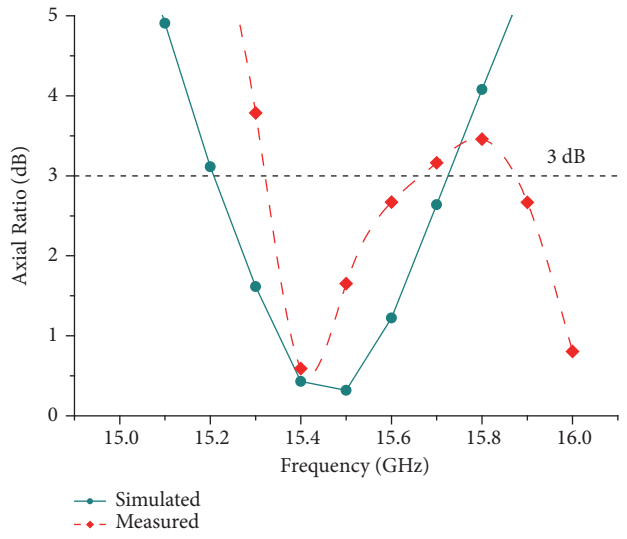


FIGURE 9: Measured axial ratio of the ASEA (simulated and measured results).

are plotted in Figure 10. In these cases, it can be seen that the proposed AESA can steer the pencil beams from -60 degrees to 60 degrees, while keeping the side lobe below about -11 dBc. It is readily observed that the measured radiated patterns are in good agreement with the simulated data. And, at both of transmitting and receiving states, the proposed ASEA keeps almost the same radiation patterns, which ensures the performance of the transceiver systems.

The antenna gain is measured including the factors such as cable loss, feeding network, and reflection loss. Figure 11 shows the simulated and measured gains at the operating frequency band. In the frequency band 15-16 GHz, the measured gain varies from 22 to 24 dBi and the average of gain is about

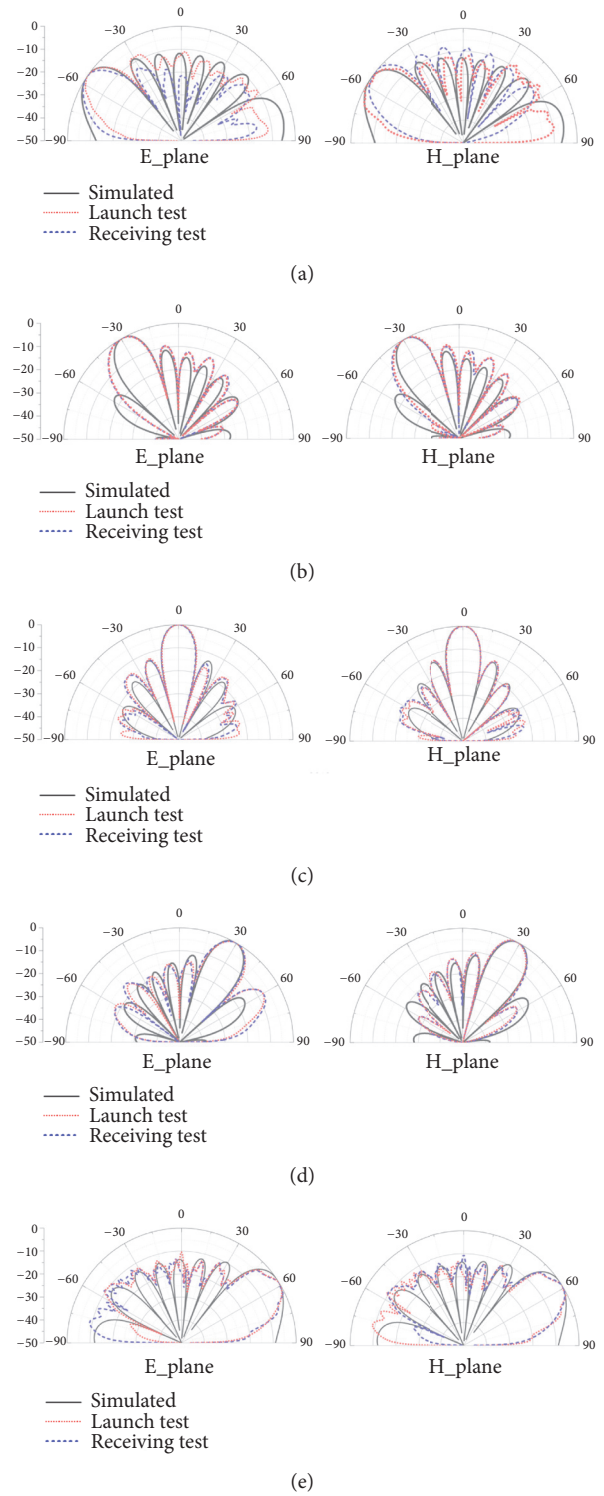


FIGURE 10: Electronically steerable 2D patterns of the proposed ASEA at 15.5 GHz: (a) AZ_-60°; (b) AZ_-30°; (c) AZ_0°; (d) AZ_30°; and (e) AZ_60°.

23 dBi. There is some slight deviation between the simulated and the measured gain, which may be due to the losses of substrate and inaccurate modeling between the simulation and fabrication.

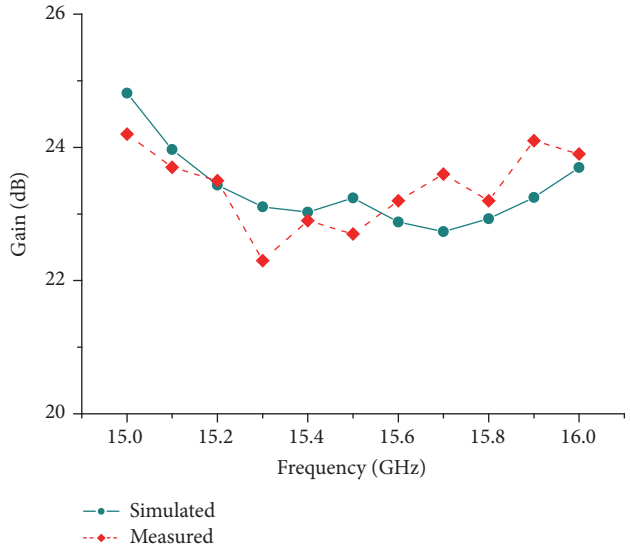


Figure 11: Gain of the proposed ASEA (simulated and measured results).

5. Conclusion

In this work, a Ku-band AESA based on 3D T/R module and LTCC patch antenna array has been introduced. By assembling two boards for different purposes into a whole multilayer board, the 3D T/R technique improved system integration and reduced implementation cost and size. Besides, a wideband circular polarized antenna array was designed in LTCC and connected to the proposed T/R modules to form a complete AESA. Measurement results showed very good phased array performances.

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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61601160 and Grant 61601161 and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY19F010013.

References

[1] J. Li, Z. He, J. Plaza et al., "Social media: new perspectives to improve remote sensing for emergency response," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1900–1912, 2017.

[2] M. Martorella, D. Pastina, F. Berizzi, and P. Lombardo, "Spaceborne radar imaging of maritime moving targets with the Cosmo-SkyMed SAR system," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 7, pp. 2797–2810, 2014.

[3] L. Corucci, A. Meta, and A. Coccia, "An X-band radar-based airborne collision avoidance system proof of concept," in *Proceedings of the 2014 15th International Radar Symposium, IRS 2014*, pp. 1-2, Poland, June 2014.

[4] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6–43, 2013.

[5] V. Nguyen, H. Nam, Y. Choe, B. Lee, and J. Park, "An X-band Bi-directional transmit/receive module for a phased array system in 65-nm CMOS," *Sensors*, vol. 18, no. 8, p. 2569, 2018.

[6] M. Kirscht, A. Dallinger, J. Mietzner, B. Bickert, J. Hippler, and R. Zahn, "Recent advances in airborne phased array radar systems," in *Proceedings of the 16th International Radar Symposium (IRS)*, pp. 231–235, 2015.

[7] N. Kundtz, K. Corp, and R. Wash, "Next generation communications for next generation satellites," *IEEE Microwave Magazine*, vol. 57, no. 8, pp. 56–64, 2014.

[8] J. Hunt, J. Gollub, T. Driscoll et al., "Metamaterial microwave holographic imaging system," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 31, no. 10, pp. 2109–2119, 2014.

[9] B. Bonnet, P. Monfraix, R. Chiniard et al., "3D packaging technology for integrated antenna front-ends," in *Proceedings of the 2008 European Microwave Integrated Circuit Conference (EuMIC)*, pp. 542–545, Amsterdam, Netherlands, October 2008.

[10] M. Hauhe and J. Wooldridge, "High density packaging of X-band active array modules," *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part B*, vol. 20, no. 3, pp. 279–291.

[11] S.-K. Yeo, J.-H. Chun, and Y.-S. Kwon, "A 3-D X-band T/R module package with an anodized aluminum multilayer substrate for phased array radar applications," *IEEE Transactions on Advanced Packaging*, vol. 33, no. 4, pp. 883–891, 2010.

[12] A. Di Carlofelice, F. De Paulis, A. Fina, U. Di Marcantonio, A. Orlandi, and P. Tognolatti, "Compact and reliable T/R module prototype for advanced space active electronically steerable antenna in 3-D LTCC technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 6, pp. 2746–2756, 2018.

[13] A. E. I. Lamminen, J. Säily, and A. R. Vimpari, "60-GHz patch antennas and arrays on LTCC with embedded-cavity substrates," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 9, pp. 2865–2874, 2008.

[14] J. Xu, Z. N. Chen, X. Qing, and W. Hong, "Bandwidth enhancement for a 60 ghz substrate integrated waveguide fed cavity array antenna on LTCC," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 3, pp. 826–832, 2011.

[15] T. Seki, N. Honma, K. Nishikawa, and K. Tsunekawa, "A 60-GHz multilayer parasitic microstrip array antenna on LTCC substrate for system-on-package," *IEEE Microwave and Wireless Components Letters*, vol. 15, no. 5, pp. 339–341, 2005.

[16] Y. Zhang, M. Sun, K. M. Chua, L. L. Wai, D. Liu, and B. P. Gaucher, "Antenna-in-Package in LTCC for 60-GHz Radio," in *Proceedings of the 2007 International workshop on Antenna Technology: Small and Smart Antennas Metamaterials and Applications*, pp. 279–282, Cambridge, UK, March 2007.

[17] D. G. Kam, D. Liu, A. Natarajan, S. Reynolds, H.-C. Chen, and B. A. Floyd, "LTCC packages with embedded phased-array

antennas for 60 GHz communications," *IEEE Microwave and Wireless Components Letters*, vol. 21, no. 3, pp. 142–144, 2011.

[18] Analysis and design of RF and digital systems using keysight SystemVue, Keysight App Note, https://literature.cdn.keysight.com/litweb/pdf/5992-0197EN.pdf.

[19] R. Caso, A. Buffi, M. Rodriguez Pino, P. Nepa, and G. Manara, "A novel dual-feed slot-coupling feeding technique for circularly polarized patch arrays," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 183–186, 2010.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 6370742, 11 pages
https://doi.org/10.1155/2019/6370742

Research Article

SUBBASE: An Authentication Scheme for Wireless Sensor Networks Based on User Biometrics

Rabia Riaz,[1] Noor-ul-Ain Gillani,[1] SanamShahla Rizvi,[2] Sana Shokat,[1] and Se Jin Kwon[3]

[1]*Department of CS & IT, University of Azad Jammu and Kashmir, Muzaffarabad 42714, Pakistan*
[2]*Raptor Interactive (Pty) Ltd., Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa*
[3]*Department of Computer Engineering, Kangwon National University, Samcheok 25806, Republic of Korea*

Correspondence should be addressed to Se Jin Kwon; sjkwon@kangwon.ac.kr

Received 9 November 2018; Revised 6 February 2019; Accepted 13 March 2019; Published 4 April 2019

Guest Editor: Fawad Zaman

To keep a network secure, a user authentication scheme that allows only authenticated users to access network services is required. However, the limited resources of sensor nodes make providing authentication a challenging task. We therefore propose a new method of security for a wireless sensor network (WSN). Our technique, Secure User Biometric Based Authentication Scheme (SUBBASe), is based on the user biometrics for WSNs. It achieves a higher security level as well as improved network performance. This solution consists of easy operations and light computations. Herein, the proposed technique is evaluated and compared with previous existing techniques. This scheme increases the performance of the network by reducing network traffic, defending against DOS attacks, and increasing the battery life of a node. Consequently, the functionality and performance of the entire network is improved.

1. Introduction

Wireless sensor networks (WSNs) contain sensor nodes for specific purposes [1]. Tiny sensor nodes in the network process information they receive after sensing their surroundings. Such networks can be applied to a large number of applications including structural health monitoring, environmental control, and battlefield surveillance [2]. In most of these applications, the user can receive data directly through a gateway node because requests are processed on this node. However, receiving data from a gateway node is occasionally difficult or even impossible. Therefore, data are obtained directly through sensor nodes [3]. Sensed data may be confidential, and illegal users can easily access sensitive data by sending a request to a sensor node. Because it is difficult for a sensor node to authenticate a request message, the leakage of sensitive information and an unnecessary depletion of network resources, e.g., node power and network bandwidth, may occur. Each of the above problems can affect the lifetime and performance of the network and make the system unattainable for legitimate users. Thus, user authentication is a required service to resist the illegal use of network data and resources [4]. To accomplish this, it is very important for sensor nodes to authenticate the identities of users. User authentication is a final solution to each of the above problems and allows authenticated users to join a network. Unfortunately it is a tremendously challenging task to provide authentication in WSNs owing to the resource limitation of its tiny sensor devices, i.e., energy and memory limitations, as well as their computational and communicational capabilities. Although various protocols have been proposed, their authentication process remains insecure [5–7]. Ultimately, to ensure the safety of a WSN, a protocol that utilizes a stronger and smarter mechanism is needed.

In this paper, we offer a competent user authentication technique for WSN applications. This scheme overcomes the authentication problem and improves the effectiveness of WSNs. We propose an authentication scheme called the Secure User Biometric Based Authentication Scheme (SUBBASe). The purposes of the designed authentication mechanism are as follows:

(i) increase the network performance by reducing network traffic;

(ii) save the battery power of the nodes, thus enhancing the lifetime of the WSN; and

(iii) defend the sensor network against different types of attacks, thus improving the functionality and performance of the entire network.

The remainder of this paper is organized as follows: Section 2 discusses the user authentication problems in existing WSN schemes through a step-by-step approach. In Section 3, we describe our proposed biometric-based user authentication protocol. Section 4 provides the security aspects of the proposed security protocol. Section 5 presents the proof of significance of SUBBASE using BAN Logic. Section 6 analyzes the performance of proposed protocol based on analytical modeling. Finally, Section 7 gives concluding remarks regarding this research.

2. Literature Review

Wong et al. [8] proposed a user authentication scheme based on the user password and cryptographic hash functions. This scheme is vulnerable to various security attacks such as forged, replay, and stolen-verifier attacks. Because a gateway and login node maintains tables containing registered user information, user passwords may be exposed by any of the sensor nodes, and a user may be blocked from altering their password [19].

Das et al. [9] proposed a user authentication scheme for a WSN that overcomes the security flaws of the scheme proposed by Wong et al. [8]. This scheme is based on the use of a smart card and the user's password. Although this scheme overcomes the weaknesses of Wong et al.'s scheme, it suffers from several security threats. For a data transmission, for example, no secure medium is provided and thus an attacker can easily alter the transmitted data. In addition, this protocol is not robust because its robustness depends on a secret parameter that is preinstalled in the sensor nodes and smart cards. If a node is captured or compromised, the security of the entire network will be harmed. In addition, an attacker can overhear entire conversation of all entities on a network. A compromised node is a major problem in this scheme and is defenseless against various types of attacks such as replaying, impersonation, password guessing, and DOS attacks.

Khan and Alghathbar [10] showed that the scheme proposed by Das et al. [9] does not provide mutual authentication and is defenseless against privileged insider attack. They determined that it is not possible to freely change a password with Das et al.'s scheme. Thus, Khan and Alghathbar proposed a security technique that attempts to overcome all of these security flaws. Using their proposed protocol, they added a user-password changing phase to Das et al.'s scheme to allow users to easily change their password. Whenever any user wishes to amend a password, the smart card overwrites the old password with a new one. To overcome the existing problems in Das et al.'s scheme, the approach proposed by Khan and Alghathbar is based on the hashed value of plain text, namely, a password. In Das et al.'s scheme, a simple password without the use of a hash value is sent to the gateway node, which causes various insider attacks to occur in a network. Thus, the hash value of the password decreases the probability of an insider attack in a network. To a certain extent, their proposed work offers security to a network by reducing the flaws in Das et al.'s scheme; however, this proposed scheme also has certain security flaws. For example, there is a problem of a session key not being established between two entities, namely, the sensor node and user, and thus mutual authentication is not provided and the problem of a lack of confidentiality of messages transmitted between participants may occur.

Yuan et al. [11] offered a protocol based on the biometrics of the user. A password and smart card are used in this particular approach. Transmitted data are not encrypted, and if an unauthorized user captures any sensor nodes, the unauthorized user can easily view the messages. Moreover, an attacker can exchange messages as a legal entity between a sensor node and user and finally gather all information available. In addition, no secure channel is provided, making message confidentiality and data integrity major problems with this particular scheme.

Yoon et al. [12] proposed an enhancement of Yuan et al.'s protocol [11] that is based on biometrics but without the use of a password. With this protocol, two secret parameters are used. Through these secret parameters, each entity verifies the legitimacy of all other entities in the network. Data integrity is considered with this protocol. However, this scheme still has security flaws because the response message of the user sent by the sensor node is not encrypted, and therefore a confidentiality problem still exists. The protocol also faces various types of DOS attacks.

To overcome the flaws to Yoon et al.'s protocol, Debiao [13] introduced another protocol based on the user's biometrics. This protocol requires complicated hardware and consumes too much time and energy. Also their protocol was vulnerable to various types of attacks, such as DOS, guessing, and replay attacks [20]. Kaul et al. [14] proposed a smart card and password-based user authentication scheme. It provides no security for user identity. It is susceptible to offline password guessing attack and smart card stolen attack and session key compromise attacks are possible. SungJin et al. [15] proposed a smart card based authentication protocol for wireless sensor network in vehicular communication.

Node compromise, message confidentiality, and data integrity are major problems with existing protocols, which also require complicated hardware. The security of previous user authentication security protocols is based on the application of a password. Short passwords are easily broken with the help of a password guessing attack. In addition, a user's password can be shared with other people or can be stolen, and there is no technique to determine an actual user. Similarly, other authentication protocols require special hardware support.

Therefore, biometric authentication is an ultimate solution to such security problems [21] and is more secure and reliable than conventional password-based authentication. Althobaiti et al. [16] pointed out numerous types of security vulnerabilities in traditional user authentication protocols

Table 1: Summary of previous authentication techniques.

Protocols	Man in the middle attack	Guessing attack	Insider attack	Replay attack	Data Integrity provided	Biometric Based	Smart Card Based
Wong [8]	Insecure	Insecure	Insecure	Insecure	No	No	No
Das and Ambani [9]	Insecure	Insecure	Insecure	Insecure	No	No	Yes
Khan and Alghathbar [10]	Insecure	Insecure	Secure	Secure	Yes	No	Yes
Yaun [11]	Insecure	Secure	Insecure	Secure	No	Yes	Yes
Yoon [12]	Insecure	Secure	Insecure	Secure	Yes	Yes	Yes
Debiao [13]	Insecure	Secure	Secure	Secure	Yes	Yes	Yes
Kaul [14]	Insecure	Insecure	Secure	Secure	No	No	Yes
SungJin [15]	Secure	Secure	Secure	Secure	Yes	No	Yes
Althobaiti [16]	Insecure	Secure	Secure	Secure	Yes	Yes	No
Dongwoo [17]	Secure	Secure	Secure	Secure	Yes	Yes	Yes
BAS [18]	Secure	Secure	Insecure	Secure	Yes	No	No

and proposed an efficient biometric-based user authentication scheme for WSNs. This scheme is feasible for resource constrained devices because it is based on a hash function and biometric encryption without the use of any complicated equipment.

Dongwoo et al. [17] removed the shortcomings of Kaul at el. [14] and proposed a more secure user authenticated key agreement method. They use user biometric based Biohash function for user authentications. Their analysis showed that their scheme is robust against all the attacks that Kaul at el. scheme was susceptible to and additionally it provides the high level of security without the requirements of time synchronization.

For authentications in sensor networks, Bi-Phase Authentication Scheme (BAS) is presented by Rabia at el. [18]. This scheme provides initial small scale authentication of the request messages entering in wireless sensor networks and provides resistance against DOS attack.

Although all of the above schemes and many other recent schemes [22–24] have suggested security improvements, there still remain drawbacks with regard to their protocols, as summarized in Table 1, such that a session key is not established after user authentication and message confidentiality is not considered. In addition, these protocols require extra hardware and are vulnerable to different types of DOS attacks. Our proposed protocol fulfills the above-mentioned shortcomings and increases the security of user authentication in a WSN.

3. Proposed User Biometric Based Secure Authentication Scheme

Owing to the exceptional characteristics of fingerprint authentication as compared to other types of biometrics, the proposed SUBBASe uses fingerprints for user authentication when joining a WSN. Moreover, a fingerprint authentication method does not require any additional hardware [25]. Users can easily provide biometric information on their own device such as a PDA or PC. To access information from the network, user can send message to sensor node directly that will be in the range of its query device. In order to query sensor node, user may use any device with fingerprint sensor, i.e., mobile phone, PDA, notebook, etc. Multiple users can be allowed to access wireless sensor network through their own mobile devices. Before network deployment, all sensor nodes are preloaded with secret information, i.e., x_0. Due to this secret information trusted node authenticates sensor node which will entertain request of user. SUBBASE considers a WSN of Mica2 sensor nodes and base station. Base station (TN) acts as authenticator of both the user and the sensor node. TN is trustworthy and secure with dominant resources in terms of memory, energy, and computation. Network architecture is shown in Figure 1. There are two main phases: an enrolment phase, followed by a user authentication phase. All symbols and notations used in this paper are described in Table 2.

3.1. How SUBBASe Algorithm Works

(1) Enrolment Phase

(i) In this phase, users register initially with a trusted node. The users then capture their biometric features and calculate a hash on them. They then submit their ID_u and hash value to the trusted node, as indicated in

$$m1 = [IDu, v] \quad \text{where } v = h(biou) \tag{1}$$

(ii) After a successful registration, trusted node computes s, as shown in (2) and sends it to the user. The value of x_o is network information applied by trusted node to extract their requested information.

$$m2 = [s] \quad \text{where } s = h(IDu \parallel xo) \tag{2}$$

Table 2: Symbols and notations.

Abbreviation	Description
ID_u	Identity of user
h(.)	One way hash function
$E_{wi}()$	Encryption of message
$D_{wi}()$	Decryption of message
‖	Concatenation operator
Δt	Time interval for transmission delay
x_o	Secret value known to TN
bio_u	Biometric of user
RI	Requested Information of user

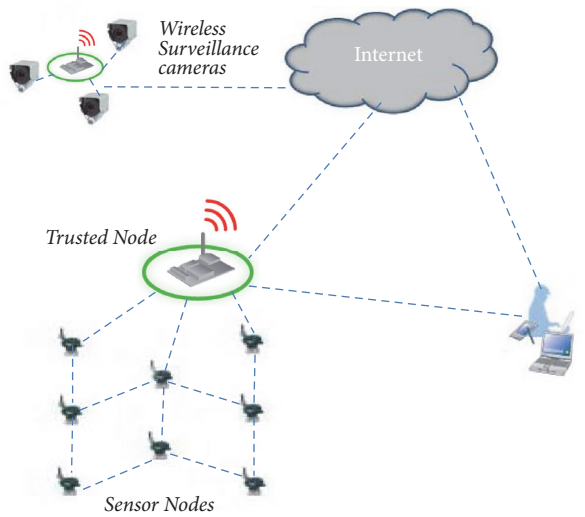


Figure 1: Depiction of wireless sensor network environment.

(2) Authentication Phase

(i) In this phase, users again capture their fresh biometric information and calculate a hash on it and then send this hash, ID_u and the requested information to the sensor node, as shown in (3). The fresh biometric information of a user is $v' = h(bio_u)$. In (3), RI is the requested information and T_0 is a user's current time stamp.

$$m3 = \left[IDu, v', RI, T0\right] \tag{3}$$

(ii) The sensor node receives a message at time T_1 and first checks the time stamp. If T_1-$T_0 \geq \Delta T$, the request is rejected; otherwise, this request is forwarded for user verification with its own *ID* to a trusted node at time T_2, as shown in (4).

Here, ΔT is the estimated time interval, and SN is the sensor node identification, which is responsible for handling user queries.

$$m4 = \left[IDu, y, T2\right] \quad \text{where } y = h\left(IDu \| v' \| SN\right) \tag{4}$$

(iii) After receiving the message at time T_3, the trusted node first checks the freshness of the message. If T_3-$T_2 \geq \Delta T$, then the request is rejected; otherwise, the trusted node checks y, as indicated in (5).

$$y = h\left(IDu \| v' \| SN\right) \tag{5}$$

(iv) The trusted node compares v and v'. If $v \neq v'$, then the trusted node sends a reject message to the sensor node.

$m_5 = [reject]$

The sensor node forwards the message to the user.

$m_6 = [reject]$

Otherwise, TN sends the message m_7 to SN.

$m_7 = [In\text{-}Progress]$

(v) When a message with a state label *In-Progress* is sent to the user, it indicates that the user can proceed to the authentication process. If a successful match occurs, the trusted node calculates *s*, as shown in (2), and

Start:
Step 1: $U \longrightarrow TN : [IDu, v]$
Step 2: $TN \longrightarrow U : [s]$
Step 3: $U \longrightarrow SN : [IDu, v', T0, RI]$
Step 4: $SN \longrightarrow TN : [IDu, y, T2]$
If
$v \neq v'$ then
exit
else
Step 5: $TN \longrightarrow SN : [in - progress]$
$TN \longrightarrow SN : [Accp, s, T4]$
Step 6: $SN \longrightarrow U[Accp, e, T6]$
Stop

Algorithm 1: Registration and authentication phase messages.

sends the message to the sensor node, as indicated in (6) at time T_4.

$$m8 = [Accp, s, T4] \tag{6}$$

(vi) The sensor node receives this message at T_5 and checks the timestamp. If T_5-$T_4 \geq \Delta T$, then the request is rejected; otherwise, *SN* computes *s*, as indicated in (2) and conducts data packing, as shown in (7) and (8).

$$d = (RI) \tag{7}$$

$$wi = h(IDu \| T6 \| s) \tag{8}$$

(vii) The sensor node uses w_i as a session key which is shared between the user and sensor node. Here, $E_{wi}(d)$ is the encryption of the requested information (RI) of the user, as indicated in (9).

$$m9 = [Accp, e, T6] \quad \text{where } e = [Ewi(d)] \tag{9}$$

(viii) The user checks the time. If T_7-$T_6 \geq \Delta T$, then the time is invalid; otherwise, it proceeds with a response. If the user is legal and holds *s*, the user will be able to access the network information. The user first computes w_i and then decrypts the message with function D_{wi}. In this way, the user obtains the requested information, as indicated in (10).

$$wi = [h(IDu \| T6 \| s)], Dwi(e) \tag{10}$$

The entire authentication algorithm of SUBBASe is summarized in Algorithm 1.

4. Security Analysis

This section demonstrates the strength of the proposed protocol from a network security perspective. We show that our proposed authentication technique has been designed to prevent various types of security attacks in WSNs, as discussed in the literature.

4.1. Stolen-Verifier Attack

If an attacker steals user information from a trusted and/or sensor node and attempts to cheat the involved entities:

this scheme can prevent a stolen-verifier attack because no password or verifier table is used. Therefore, an attacker cannot steal user information from a trusted and/or sensor node. Schemes that preserve such password tables to confirm a user login may suffer from this type of attack. This threat is solved with SUBBASe; however, because user biometrics is used for the network login.

4.2. Message Confidentiality

If an attacker overhears a message exchanged between a legal user and a sensor node and obtains secret network information:

with SUBBASe, message confidentiality is provided because user requested information is encrypted. If requested information is transmitted without encryption and passes through a public channel, the attacker can easily sniff the network information. Existing protocols do not provide message confidentiality. SUBBASe provides this service as encrypted requested information, $E_{sk}(d)$, is sent across a network.

4.3. Provide Mutual Authentication

If both parties are not authenticated during conversation:

SUBBASe provides mutual authentication because all entities authenticate each other. For example, when a user sends a message to *SN*, *SN* then verifies it by sending a request to *TN*. In addition, *TN* verifies *SN* based on its *ID*, and when *TN* sends a message to the user, it verifies it through hidden parameter x_0.

4.4. Complicated Equipment

If there is requirement of additional hardware for the authentication process:

various storage devices such as smart cards are used for identification purposes, providing better security. However, they require special and expensive hardware, which not everyone can afford. With SUBBASe, users can easily use any recognition interface without the need for an additional device using its Mobile or PDA. This is beneficial to both users and vendors and can increase the network efficiency and user convenience. Through the development of new technologies, mobile or laptop devices can easily be used to identify numerous types of biometrics.

4.5. Guessing Attack

If an attacker is able to guess the password or security parameters used for authentication purpose:

SUBBASe offers resistance to a guessing attack because users provide biometric features at the time of registration and at the time of authentication. Thus, an attacker cannot guess the user's biometrics. Preventing a guessing attack is crucial in systems that are based on password security.

4.6. Data Integrity

If an attacker obtains a message transmitted between a sensor node and a legal user and makes changes to the content:

this service provides assurance that communicated data cannot be changed by an unauthorized entity. With SUBBASe, data integrity is provided using a one-way hash function, i.e., $m_4 = [ID_u, y, T_2]$, where $y = h(ID_u \| v' \| SN)$, which is a message sent by the user to *TN*. Similarly, all communicated messages are sent in the same way, which cannot be modified by an attacker.

4.7. Prevent Replay Attack

If an attacker obtains previously communicated messages and starts communicating after acquiring the same rights as a legal user:

SUBBASe uses a time stamp to prevent this type of an attack. Suppose an attacker captures $m_3 = [ID_u, RI, v', T_0]$ and wants to replay the same message to *TN*. If the attacker perceives the communication message, the attacker cannot get verify because T_1-$T_0 \geq \Delta T$, where T_1 is the time stamp when the replayed message is received by *TN*. Time synchronization for wireless sensor networks is a very active research area [26–28] and it can easily be used for prevention of replay attack.

4.8. Prevent Node Compromise Attack

If an attacker succeeds in capturing a sensor node and collects all sensitive data from it:

if user is allowed to get data directly from sensor node without authenticating the node, it will result in the attack "node compromise". In SUBBASE sensor node is first authenticated by trusted node after authentication; sensor node is able to respond to the query of user that prevents node compromise attack. Also user hash of biometric is saved on the node which cannot be retrieved, as hash is a one-way function.

4.9. Network Traffic Attack

If an attacker succeeds in sending too much traffic on network like DDos attacks, with the intention to disrupt network functionality:

in SUBBASE protocol, authentication messages are reduced as task is divided among TN and SN to authenticate users. TN authenticates SN and user simultaneously, in order to avoid other malicious attacks like Denial of Service (DOS) attack. For example, if SN is not authenticated by TN, then any malicious node can send multiple fake authentication request messages/packets over the network that will results in increasing the network's traffic, depleting the resources of TN, and denying services to original users.

Table 3 provides an enhanced comparison of the different types of previous schemes against SUBBASe with respect to message confidentiality, network attacks, security parameters, and the requirement of any additional hardware devices for authentication.

5. Proof of SUBBASE Using BAN Logic

In this section we will use Burrows-Abadi-Needham (BAN) logic to validate that user and sensor node generate a valid and fresh session key for message exchange. Basic symbols used for BAN logic are described in Table 4.

The ban logic also provides the following basic rules:

(1) Message meaning rule:

$$\frac{U \mid\equiv U \overset{K}{\longleftrightarrow} S, U \lhd \{X\} K}{U \mid\equiv S \mid\sim X} \tag{11}$$

(2) Nonce verification rule:

$$\frac{U \mid\equiv \#(X), U \mid\equiv S \mid\sim X\ )}{U \mid\equiv S \mid\equiv X} \tag{12}$$

(3) The believe rule:

$$\frac{U \mid\equiv (X, Y)}{U \mid\equiv X} \tag{13}$$

Table 3: Comparison between SUBBASe and existing schemes.

Attacks	Debiao	Yoon	Althobaiti	Kaul	Dongwoo	SUBBASe
Session key Establishment	No	No	Yes	Yes	Yes	Yes
Message Confidentiality	No	No	Yes	Yes	Yes	Yes
Prevent Integrity	Yes	Yes	Yes	No	Yes	Yes
Prevent Replay Attack	Yes	Yes	Yes	Yes	Yes	Yes
Prevent Guessing Attack	Yes	Yes	Yes	No	Yes	Yes
Provide Mutual Authentication	Yes	Yes	Yes	Yes	Yes	Yes
Avoid Impersonation attack	No	No	No	No	Yes	Yes
Avoid Node Compromise Attack	No	No	Yes	No	Yes	Yes
Complicated Hardware Needed	Yes	Yes	No	Yes	Yes	No
Avoid Insider Attack	Yes	No	Yes	Yes	Yes	No

Table 4: Symbols for BAN Logic.

Symbol	Meaning
U $\mid\equiv$ X	U *Believe* the Statement X
#X	The Statement X is *Fresh*
U $\mid\sim$ X	U once *Said* X
$\{X\}_K$	Formula X is *encrypted* by key K
$\langle X\rangle_K$	Formula X is *combined* by key K
U $\Longrightarrow$ X	U *Control over* the Statement X
U $\triangleleft$ X	U *See* the Statement X
U $\overset{K}{\longleftrightarrow}$ S	U and S shared key K for communication
SK	Session Key
T	Time Stamp

(4) Freshness rule:

$$\frac{U \mid\equiv \#(X)}{U \mid\equiv \#(X,Y)} \tag{14}$$

(5) Jurisdiction rule:

$$\frac{U \mid\equiv S \mid\Longrightarrow X, U \mid\equiv S \mid\equiv X}{U \mid\equiv X} \tag{15}$$

We have the following goals to prove the validity of requested information and freshness of session key used for communication:

Goal 1. $U \mid\equiv \#(SK)$

Goal 2. $U \mid\equiv \#(RI)$

The message exchange of SUBBASE in idealized form is given below:

Message 1. $U \longrightarrow SN : \langle IDu, v', RI\rangle$

Message 2. $SN \longrightarrow TN : \langle IDu, Y\rangle$

Message 3. $TN \longrightarrow SN : (s)$

Message 4. $SN \longrightarrow U : T, \{RI\}sk$

To proceed with the proof, we have defined the following assumptions:

(i) A1: $U \mid\equiv SN \Longrightarrow SK$

(ii) A2: $SN \mid\equiv TN \overset{X_O}{\longleftrightarrow} SN$

(iii) A3: $TN \mid\equiv TN \overset{X_O}{\longleftrightarrow} SN$

(iv) A4: $SN \mid\equiv SN \overset{SK}{\longleftrightarrow} U$

(v) A5: $U \mid\equiv SN \mid\sim T$

(vi) A6: $U \mid\equiv TN \mid\sim s$

(vii) A7: $U \mid\equiv \#(RI)$

Detailed Process of Proof Is as Follows

Step 1. According to M4

$$V1 : SN \triangleleft T, \{RI\}\, SK \tag{16}$$

Table 5: Comparison of computational time between SUBBASe and existing schemes.

Protocols	Registration	Login+ Authentication	Total	Total time
Yoon	3 T_H	10 T_H	13 T_H	47.26 ms
Debiao	3 T_H	8 T_H +4t_{sym}	11 T_H +4t_{sys}=15 T_H	54.54 ms
Althobaiti	2 T_H	6 T_H + 2T_{MAC} +4 RC_5	8T_H + 2T_{MAC} +4 RC_5	36.37ms
Kaul	6T_H	16T_H	22T_H	79.97ms
Dongwoo	6T_H	14 T_H	20 T_H	72.7ms
SUBBASe	2T_H	7T_H +2Trc5	9 T_H +2*T*rc5	33.24 ms

Step 2. According to V1 and A4 and message meaning rule

$$V2 : U \mid\equiv SN \mid\sim RI \tag{17}$$

Step 3. According to A7 and V2 and nonce verification rule

$$V3 : U \mid\equiv SN \mid\equiv RI \tag{18}$$

Step 4. According to V3, A1 and believe rule

$$V4 : U \mid\equiv (RI, SK) \tag{19}$$

Step 5. According to A7 and V4 and freshness rule

$$\begin{aligned} V5 : U \mid\equiv \#(SK) \quad (GOAL\ 1) \\ V6 : U \mid\equiv \#(RI) \quad (GOAL\ 2) \end{aligned} \tag{20}$$

6. Performance Analysis

We use a mathematical model to examine the performance of SUBBASe and how it compares with existing schemes. The comparison is based on a computation of the time and energy consumption. We selected RC5 which is most suitable for implementation on resource constrained devices. The time to encrypt and decrypt a message using RC5 on a mica2 node is 0.26ms [29]. Similarly we chose SHA-1, whose performance time for a one-way hash function on mica2 is 3.636ms [30].

With Yoon et al.'s scheme 13 hash computation operations are required. A user needs one hash operation during the registration phase, and three during the authentication phase. A sensor node needs three hash calculations, and a trusted node needs two hash operations during the registration phase and four during the authentication phase. Similarly, the Debiao protocol [13] requires 11 hash computation operations and four T_{sym} operations for calculating a symmetric function (encryption/decryption) during the login and authentication phase. A user needs one hash function during the registration phase, and three hash operations and one T_{sym} operation for the encryption/decryption function during the authentication phase. A sensor node needs two hash calculations and one T_{sym} encryption/decryption function, and a trusted node requires two hash operations during the registration phase, and three hash operations and two T_{sym} operations during the authentication phase. A T_{sym} operation has the same computational cost as a hash operation, and thus the total number of hash operations with the Debiao protocol is 15. In the same manner, Althobaiti et al.'s scheme [15] requires two hash functions during the registration phase and six hash functions, i.e., two MAC functions and four RC5 functions, during the login and authentication phase. In Kaul et al.'s scheme user needs to perform 2 hash operations in registration phase while trusted node has to perform 4 hash operations. In login phase, user smart card performs 8 hash operations. During authentication, trusted node performs 6 and user performs 2 hash operations. Password change phase requires 10 hash operations by smart card. In Dongwoo et al.'s scheme user and trusted node both need to perform 3 hash operations in registration. In login phase, user side smart card performs 6 hash operations. During authentication, trusted node performs 5 and user performs 3 hash operations. Password change phase requires 7 hash operations by smart card.

With SUBBASe, one hash operation is required by a user during the enrolment phase; two hash operations and time for message decryption and three hash operations and time for message encryption are required by a sensor node; and three hash operations are required by a trusted node. Because a tiny sensor node has an inadequate amount of energy, the aim of our protocol is to reduce the computational cost of a sensor node. Although a user and a trusted node have sufficient resources to conduct multiple tasks, our scheme also minimizes the computational cost of a trusted node. Table 5 shows a complete picture of our comparison.

We used Matlab simulation to evaluate the strength and performance of the proposed security technique. Figure 2 shows a time computation graph of SUBBASe, and the Yoon et al., Debiao, Althobaiti et al., Kaul et al., and Dongwoo et al. schemes for multiple users accessing a network at the same time. When the number of users increases, the computational overhead for authentication also increases because more nodes are involved in the authentication process. With the SUBBASe scheme, when there is only one user, the computational overhead is 33.24ms, whereas with Yoon et al.'s scheme it is 47.26ms, in Debiao's method it is 54.54ms, with Althobaiti et al.'s approach it is 36.37ms, and with Kaul et al. and Dongwoo et al. it is 79.97ms and 72.7ms respectively. Results show that SUBBASe improves the network performance and its lifetime by reducing the amount of overhead.

Similarly, we calculated the energy consumption of SUBBASe and the other existing protocols. Main contributor for energy consumption in wireless sensor networks is data

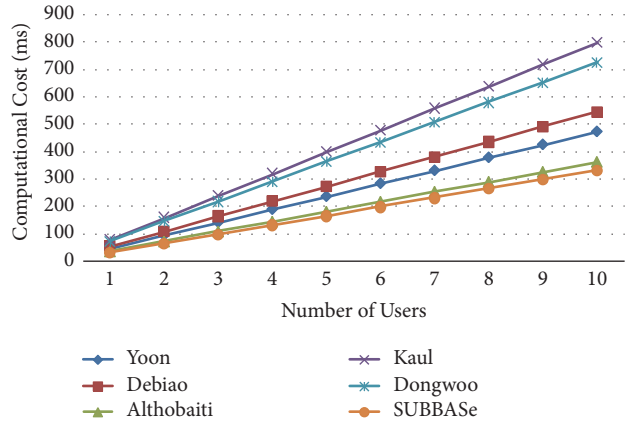


Figure 2: Computational overhead with respect to number of users.

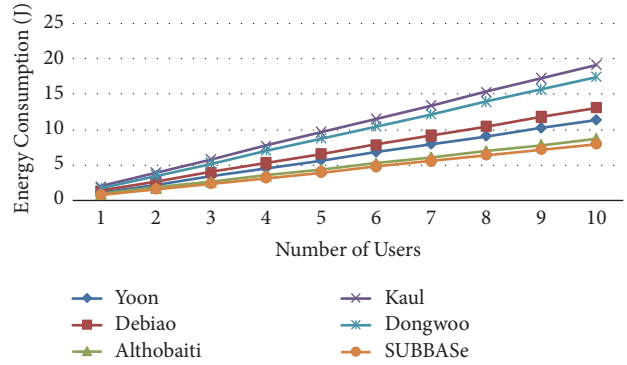


Figure 3: Energy consumption with respect to number of users.

transmission and reception, i.e., power consumed by antenna for sending and receiving messages. Just for comparison purpose, this paper focuses on calculating how much node power is consumed while performing the computations for authentication function.

To calculate the energy, we used "Computational Energy Cost" equation described in [31] and neglected the inactive state component. During the authentication process, sensor nodes do not go into inactive state. For mica2, node v = 3.0, and I = 8mA. In the equation, E denotes the consumption of energy, Q is the charge, I is the current, V is the voltage, and t is the time.

$$E = V \; x \; I \; x \; t \tag{21}$$

Figure 3 shows the energy consumption of SUBBASe, and the Yoon, Debiao, Althobaiti et al., Kaul et al., and Dongwoo et al. schemes with respect to the number of users. With SUBBASe, for one user, the energy consumption is 0.79J, whereas for Yoon et al.'s scheme it is 1.13J, for Debiao's approach it is 1.30J, for Althobaiti et al.'s method it is 0.87J, and for Kaul et al. and Dongwoo et al. schemes it is 1.9J and 1.7J respectively. Thus, the energy consumption of our proposed scheme is much less than that of the other existing security protocols. SUBBASe therefore proves to be more efficient, increasing the network performance and the lifetime by saving battery power of the nodes.

7. Conclusion and Future work

We proposed an authentication protocol that is based on the biometrics of users without the use of any traditional password or extra hardware devices. The proposed protocol simply proves the identity of the users through their biometrics. In addition, we designed our protocol to use simple and light computations. We mathematically analyzed the performance and security capability of SUBBASe and proved that it has better security features than other existing approaches. Moreover, based on a comparison with existing protocols, its computational cost and energy consumption are deemed to be suitable for resource constrained networks. In future BAN logic can be applied on SUBBASE to check its freshness property and its simulation can be done in security analysis tools.

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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A3B04031440). This study was also supported by 2017 Research Grant from Kangwon National University.

References

[1] C. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–105, 2002.

[3] Y. Faye, I. Niang, and T. Noel, "A survey of access control schemes in wireless sensor networks," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 5, no. 11, 2011.

[4] H. Wang and Q. Li, "Achieving distributed user access control in sensor networks," *Ad Hoc Networks*, vol. 10, no. 3, pp. 272–283, 2012.

[5] P. Shantala, K. Vijaya, S. Sonali, and J. Rashique, "A survey on authentication techniques for wireless sensor networks," *Int. Journal of Applied Engineering Research*, vol. 7, no. 11, 2012.

[6] R. Rasmita and S. Itun, "A survey on authentication protocols for wireless sensor networks," *Int. Journal of Engineering Science and Technology*, vol. 3, no. 5, pp. 4253–4256, 2011.

[7] C. Jiang, C. B. Li, and H. Xu, "An efficient scheme for user authentication in wireless sensor networks," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops/Symposia (AINAW '07)*, vol. 1, pp. 438–442, Ontario, Canada, May 2007.

[8] K. H. M. Wong, Z. Yuan, C. Jiannong, and W. Shengwei, "A Dynamic User Authentication Scheme for Wireless Sensor Networks," in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, vol. 1, pp. 244–251, IEEE, Taichung, Taiwan, June 2006.

[9] M. L. Das, "Two-factor user authentication in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1086–1090, 2009.

[10] M. K. Khan and K. Alghathbar, "Cryptanalysis and security improvements of 'two-factor user authentication in wireless sensor networks'," *Sensors*, vol. 10, no. 3, pp. 2450–2459, 2010.

[11] J. Yuan, C. Jiang, and Z. Jiang, "A biometric-based user authentication for wireless sensor networks," *Wuhan University Journal of Natural Sciences*, vol. 15, no. 3, pp. 272–276, 2010.

[12] E.-J. Yoon and K.-Y. Yoo, "A new biometric-based user authentication scheme without using password for wireless sensor networks," in *Proceedings of the 2011 20th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE 2011*, pp. 279–284, France, June 2011.

[13] H. Debiao, "Robust biometric-based user authentication scheme for wireless sensor networks," *Cryptology ePrint Archive*, vol. 203, pp. 1–15, 2012.

[14] S. D. Kaul and A. K. Awasthi, "Security Enhancement of an Improved Remote User Authentication Scheme with Key Agreement," *Wireless Personal Communications*, pp. 1–17, 2016.

[15] S. Yu, J. Lee, K. Lee, K. Park, and Y. Park, "Secure Authentication Protocol for Wireless Sensor Networks in Vehicular Communications," *Sensors*, vol. 18, no. 10, 2018.

[16] O. Althobaiti, R. Mznah, and A. Abdullah, "An efficient biometric authentication protocol for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 5, 2013.

[17] D. Kang, J. Jung, H. Kim, Y. Lee, and D. Won, "Efficient and secure biometric-based user authenticated key agreement scheme with anonymity," *Security and Communication Networks*, vol. 2018, Article ID 9046064, 14 pages, 2018.

[18] R. Riaz, T.-S. Chung, S. S. Rizvi, and N. Yaqub, "BAS: the biphase authentication scheme for wireless sensor networks," *Security and Communication Networks*, vol. 2017, Article ID 7041381, 10 pages, 2017.

[19] M. Rakesh, *An improved user authentication protocol for hierarchical wireless sensor networks using elliptic curve cryptography [M. S. thesis]*, Department of Computer Science and Engineering, 2012.

[20] G. Padmavathi and D. Shanmugapriya, "A survey of attacks, security mechanisms and challenges in wireless sensor networks," *International Journal of Computer Science and Information Security*, vol. 4, no. 1, 2009.

[21] B. Debnath, R. Rahul, A. Farkhod, and C. Minkyu, "Biometric authentication: A review," *Science and Technology*, vol. 2, no. 3, 2009.

[22] W. Yang, J. Hu, S. Wang, and Q. Wu, "Biometrics based privacy-preserving authentication and mobile template protection," *Wireless Communications and Mobile Computing*, Article ID 7107295, 17 pages, 2018.

[23] Z. Han, L. Yang, S. Wang, S. Mu, and Q. Liu, "Efficient multifactor two-server authenticated scheme under mobile cloud computing," *Wireless Communications and Mobile Computing*, vol. 2018, 14 pages, 2018.

[24] E. Pagnin and A. Mitrokotsa, "Privacy-preserving biometric authentication: challenges and directions," *Security and Communication Networks*, vol. 2017, Article ID 7129505, 9 pages, 2017.

[25] B. Roli, S. Priti, and B. Punam, "Minutiae extraction from fingerprint images - a review," *International Journal of Computer Science Issues*, vol. 8, no. 5, 2011.

[26] D. Upadhyay, A. K. Dubey, and P. Santhi Thilagam, "Time synchronization problem of wireless sensor network using maximum probability theory," *International Journal of Systems Assurance Engineering and Management*, vol. 9, no. 2, pp. 517–524, 2018.

[27] L. Liu, G. Luo, K. Qin, and X. Zhang, "An on-demand global time synchronization based on data analysis for wireless sensor networks," *Procedia Computer Science*, vol. 129, pp. 503–510, 2018.

[28] N. Xiong, M. Fei, T. Yang, and Y. Tian, "Randomized and efficient time synchronization in dynamic wireless sensor networks: a gossip-consensus-based approach," *Complexity*, vol. 2018, 16 pages, 2018.

[29] X. Le, M. Khalid, R. Sankar, and S. Lee, "An efficient mutual authentication and access control scheme for wireless sensor networks in healthcare," *Journal of Networks*, vol. 6, no. 3, pp. 355–364, 2011.

[30] C. Karlof, N. Sastry, and D. Wagner, "TinySec: a link layer security architecture for wireless sensor networks," in *Proceedings of the Second International Conference on Embedded Networked Sensor Systems (SenSys '04)*, pp. 162–175, 2004.

[31] M. A. Razzaque and S. Dobson, "Energy-efficient sensing in wireless sensor networks using compressed sensing," *Sensors*, vol. 14, no. 2, pp. 2822–2859, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 8186573, 14 pages
https://doi.org/10.1155/2019/8186573

Research Article

Enabling Noninvasive Physical Assault Monitoring in Smart School with Commercial Wi-Fi Devices

Qizhen Zhou,[1] **Chenshu Wu**,[2] **Jianchun Xing**,[1] **Shuo Zhao**,[1] **and Qiliang Yang**[1]

[1]*National Defence Engineering College, Army Engineering University of PLA, Nanjing 210007, China*
[2]*Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA*

Correspondence should be addressed to Qiliang Yang; yql@893.com.cn

Received 8 November 2018; Revised 26 February 2019; Accepted 14 March 2019; Published 1 April 2019

Guest Editor: Sungchang Lee

Monitoring physical assault is critical for the prevention of juvenile delinquency and promotion of school harmony. A large portion of assault events, particularly school violence among teenagers, usually happen at indoor secluded places. Pioneering approaches employ always-on-body sensors or cameras in the limited surveillance area, which are privacy-invasive and cannot provide ubiquitous assault monitoring. In this paper, we present Wi-Dog, a noninvasive physical assault monitoring scheme that enables privacy-preserving monitoring in ubiquitous circumstances. Wi-Dog is based on widely deployed commodity Wi-Fi infrastructures. The key intuition is that Wi-Fi signals are easily distorted by human motions, and motion-induced signals could convey informative characteristics, such as intensity, regularity, and continuity. Specifically, to explicitly reveal the substantive properties of physical assault, we innovatively propose a set of signal processing methods for informative components extraction by selecting sensitive antenna pairs and subcarriers. Then a novel signal-complexity-based segmentation method is developed as a location-independent indicator to monitor targeted movement transitions. Finally, holistic analysis is employed based on domain knowledge, and we distinguish the violence process from both local and global perspective using time-frequency features. We implement Wi-Dog on commercial Wi-Fi devices and evaluate it in real indoor environments. Experimental results demonstrate the effectiveness of Wi-Dog which consistently outperforms the advanced abnormal detection methods with a higher true detection rate of 94% and a lower false alarm rate of 8%.

1. Introduction

School violence, as the leading cause of juvenile delinquency, has become an increasingly serious social issue and attracted extensive academic attention from researchers. According to a report of the National Center for Education Statistics, 28% of total 4326 examined adolescents reported bullying victimization, whose physical and mental health were severely affected [1]. To curb the prevalence of school violence, governments have introduced relevant policies to deal with it. A key enabler for effective school violence prevention is to automatically detect and alarm the instantaneous physical assault with existing available infrastructures. Wearable sensor based scheme may provide possible approaches to monitor a specific group of users, especially the guarded ones [2]. However, the always-on-body dedicated sensors (e.g., data glove [3], RFID [4], and smartphone [5]) render it not only uncomfortable to comply with but also inapplicable to general-purposed monitoring. More common solutions resort to camera-based monitoring [6–8]. Premounted cameras continuously collect and analyze the video frames of areas-of-interests, yet they bring underlying privacy issues and only operate in a clear line-of-sight (LOS) view.

Recent innovations in the increasingly hot area of wireless human sensing [9–11] inspire us to develop an upgraded assault monitoring system using Wi-Fi signals. Similar to many other activity recognition systems, the rationales of Wi-Fi-based assault detection are twofold. On the one hand, Wi-Fi signal is pervasive nowadays with densely deployed Wi-Fi infrastructures in public places, which delivers the idea of ubiquitous device-free surveillance into a practical solution. On the other hand, abrupt physical assault along with rapid movements of body parts could alter Wi-Fi signals and thus encode distinct features in received signals. Such features can be effectively captured by PHY layer Channel State Information (CSI), which is measurable on commercial Wi-Fi

devices. Instead of the traditional Received Signal Strength (RSS), CSI conveys subcarrier-level channel measurements, with natural advantages in revealing the characteristics of superposed signals, and has been widely used for wireless sensing.

Emerging CSI-based technologies have promoted the revolution of action recognition domain, covering from macro-movements to micro-movements [12–20]. Existing approaches, however, cannot be directly applied to assault monitoring since previous works usually rely on the repetitive patterns of actions [14, 17] or distinct profiles of single-person specialized gestures [18–20]. In contrast, assault events are likely irregular and unpredictable. Compared with normal single-user activities, physical assault events are differentiated based on three criteria [5]. (1) *High-intensity.* Multiple users (both attackers and victims) in the process of physical conflicts behave aggressively, inducing rapid and intensive body movements. (2) *Irregularity.* In a real situation, physical assault cannot be regarded as repetitions of simple actions. Instead, the assault-induced signals generate complex and disordered fluctuations with the escalation of physical assault. (3) *Continuity.* Severe physical assault always happens to specific victims, while the procedure of abuse can last for a long time. With in-depth understanding of the violent process and elaborate analysis of the event properties, it seems possible to recognize a violent event from RF signals.

In this paper, we show for the first time the feasibility of leveraging Wi-Fi signals for monitoring physical assault. We present Wi-Dog, a noninvasive physical assault monitoring scheme on commercial Wi-Fi infrastructures, to protect users from potential physical assault, just like a loyal dog does. Wi-Dog enables a privacy-preserving manner for daily assault monitoring in ubiquitous environments and advances the state-of-the-art Wi-Fi-based sensing approaches by solving three critical challenges. (1) *How to obtain abundant motion information from noisy CSI dynamics?* Assault-induced fluctuations in CSI is easily distorted and blurred by background noises and irrelevant body movements. To reveal the genuine CSI waveforms, some previous works extracted the first principle component which still suffered from the noise interferences [16] or the second principle component which lost the majority of motion information [15, 18]. Instead, Wi-Dog obtains abundant and accurate CSI fluctuations by taking advantage of spatial diversity. The key intuition is that the same subcarrier of different antennas suffers from various channel distortions but the same noise sources, generating some similar variations in waveforms. Therefore, we propose a series of noise reduction steps to properly manipulate CSI dynamics from multiple subcarriers, eliminating irrelevant interferences while retaining motion cues of interests to the greatest extent. (2) *How to precisely and sensitively detect abnormal transitions during human interactions?* As the drastic conflicts are location-variant and complex, conventional variance-based segmentation [21] cannot be utilized because slight interactions near the links may induce higher variation in CSI waveforms rather than assault. To address this challenge, we resort to the signal complexity of target frequency band to monitor the variation of intensity and irregularity level. The cross correlation of adjacent subcarriers is also supplemented to enhance the capability of location-independent detection. (3) *How to differentiate real assaults from assault-like actions?* Existing fall-like detection works [21] fully utilized the features extracted from both amplitude and phase information in time-frequency analysis, which cannot be directly applied in our amplitude-based approach. Moreover, based on previous experimental study [5, 22], we notice that assault process can be easily mistaken as strenuous exercise (e.g., run, frog leap, and exergame) or normal human interactions (e.g., talk with body language). To make a comprehensive identification of assault process, we firstly extract novel features representing intensity level from Doppler frequency shifts and adopt a SVM classifier to classify time segments with high-intensity features, which is termed as local analysis. To reduce the false alarm rate, we further consider the irregularity as well as continuity in longer time duration by counting the occurrence frequency of new patterns and underlying slices.

We prototype Wi-Dog with commercial Wi-Fi devices and validate its performance in real environments. Experimental results show that Wi-Dog can monitor the imitated physical attacks with a high true detection rate of 0.94 (0.85), along with a low false alarm rate of 0.08 (0.11), using only a single pair of Wi-Fi transmitter-receiver in LOS (NLOS) environment. The results also show that Wi-Dog is robust to rational changes of parameters, including thresholds, individual diversity, duration time, and sampling rate. In a nutshell, our contributions are summarized as follows.

(i) To the best of our knowledge, Wi-Dog is the first work to present a noninvasive physical assault monitoring system with only a pair of commercial Wi-Fi devices. We empower pervasive available Wi-Fi signals with the sensing ability and expand the boundaries of Wi-Fi to a new realm. We envision that this capability could enact as an early step toward more general emergency detection applications, including, but not limited to, terrorist threat warning, elders' healthcare, and injury rescue.

(ii) We drill down the domain of human abnormal interactions and explore the feasibility of Wi-Fi-based physical assault monitoring. The key enabler is to fully exploit the high-intensity, irregularity, and continuity of human motions reflected in CSI measurements. Hence we conduct holistic analysis and classify the violent events from both local and global perspective using elaborate-designed features.

(iii) We innovatively recover the informative motion-induced signals from noisy CSI measurements and design a location-independent indicator to detect the physical assault based on signal complexity.

(iv) We perform extensive experiments and validate Wi-Dog in both classroom (NLOS) and corridor environments (LOS) by imitating real physical assault with different volunteers. Experimental results show that Wi-Dog outperforms the state-of-the-art abnormal detection approaches in assault monitoring with a high true detection rate of 0.94 (0.85), along with a

low false alarm rate of 0.08 (0.11) in LOS (NLOS) environment with rational changes of parameters.

2. Preliminary and Background

In this section, we present the critical intuitions of assault monitoring system and introduce the concept of Doppler Effect in CSI.

2.1. Analysis of Physical Assault. Formally, physical assault is properly defined by the World Health Organization (WHO) as *the intentional use of physical power toward another person which may result in deadly injuries* [23]. The definition involves intentions and outcomes of assault itself, irrespective of the essential characteristics it contains. To understand and explore the assault process, we resort to Violent-Flows Database [24], which is publicly recognized as an evaluation benchmark for crowd assault detection. All 246 videos (123 assault and 123 non-assault) are downloaded from YouTube, with an average 3.6s video clip. Observing uncontrolled assault conditions in real world, we believe that a practical assault monitoring systems should meet the following requirements. (1) *Privacy-preserving*. A monitoring system should not violate the privacy of users, especially the normal users presenting in the surveillance area. (2) *Device-free*. One should never expect an attacker to expose himself and wear any devices to be monitored. In contrast, an attacker would typically seek ways to hide from any monitoring system. Hence a useful system should work in passive and noncontact mode, allowing monitoring of attackers without any devices. (3) *Omnidirectional*. Assault events frequently happen at blind corners that are not easy to be covered by traditional monitoring like camera-based systems. Ubiquitous monitoring systems should provide omnidirectional coverage for as large area as possible. Inspired by recent innovations on CSI-based human behaviour sensing, we believe the ubiquitously existing Wi-Fi signal provides an alternative promising way to monitor indoor assault, which fulfils all the above conditions.

2.2. Doppler Effect in CSI. The proof-of-concept assault monitoring system highly relies on the extraction of Doppler shift in CSI. Doppler shift is described as the change in wavelength of a signal for receivers, which results from relative motions of target objects, transmitters, and receivers. In the context of wireless sensing, if we consider a moving human reflecting the multipath signals as a virtual transmitter [25], then a human performing full-body motions could generate distinct Doppler shift in receivers. Therefore, the motion-induced variation of k-th path in frequency is given by

$$f_{D_k} = -\frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{d}x} d_k(x) \tag{1}$$

where f_{D_k} is the Doppler frequency shift for the k-th path, λ denotes the wavelength of the signal in the medium, and $d_k(\mathrm{x})$ implies the length of the k-th propagation path. That is, the quicker the motion is performed, the faster the path length changes and the larger the Doppler shift is produced. To further depict channel properties of multiple paths, we introduce the concept of CSI [26], the superimposed channel response of each individual path, which can be written as

$$H(f,t) = e^{-j2\pi\Delta f t} \sum_{n=1}^{N} \partial_k(\mathrm{t})\, e^{-j2\pi f \tau_k(\mathrm{t})} \tag{2}$$

where $e^{-j2\pi\Delta f t}$ implies the phase shift caused by carrier frequency offset (CFO), Δf is the carrier frequency, N is the total number of propagation paths, and $\partial_k(\mathrm{t})$ denotes the attenuation factor for the k-th path at time t. Considering that $d_k(\mathrm{x})$ can be expressed as the product of the speed of light c and time of flight $\tau_k(\mathrm{t})$, another expression of $\tau_k(\mathrm{t})$ is $\tau_k(\mathrm{t}) = \mathrm{d}_k(\mathrm{t})/c = (1/f)\int_{-\infty}^{t} f_{D_k}(\mathrm{x})\mathrm{dx}$. To understand the relationship between CSI and Doppler Effect, we slit CSI into static component $H_s(f)$ ($f_D = 0$) and dynamic component $H_d(f)$ ($f_D \neq 0$). As human motion may change a set of path length, $H_d(f)$ can be unfolded as:

$$H_d(f) = \sum_{k\in P_d} \partial_k(t)\, e^{j2\pi \int_{-\infty}^{t} f_{D_k}(x)\mathrm{d}x} \tag{3}$$

where P_d denotes the set of dynamic path. To eliminate phase noises and CFO, the unwrapped instantaneous CSI power is calculated as follows.

$$\begin{aligned} \left|H(f,t)\right|^2 = {} & \sum_{k\in P_d} 2\left|H_s(f)\,\partial_k(t)\right| \\ & \cdot \cos\left(2\pi \int_{-\infty}^{t} f_{D_k}(x)\,\mathrm{d}x + \varnothing_{sk}\right) \\ & + \sum_{k,l\in P_d} 2\left|\partial_k(t)\,\partial_l(t)\right| \\ & \cdot \cos\left(2\pi \int_{-\infty}^{t} \left(f_{D_k}(x) - f_{D_l}(x)\right)\mathrm{d}x + \varnothing_{kl}\right) \\ & + \sum_{k\in P_d} \left|\partial_k(t)\right|^2 + \left|H_s(f)\right|^2 \end{aligned} \tag{4}$$

Noticing that the frequency of sinusoids can be extracted, we use Hilbert Transform to evaluate the speeds of human body parts for assault analysis. In the following sections, we would introduce the system architecture and design of Wi-Dog.

3. Overview

Wi-Dog is a device-free physical assault monitoring system using only a pair of commercial Wi-Fi devices. Figure 1 shows the architecture overview of Wi-Dog, which consists of four critical components, i.e., CSI collection, processing, segmentation, and assault recognition step. Specifically, we resort to fine-grained CSI tractable on commercial NICs to explore the underlying characteristics of assault-induced variations. The rationale is that furious physical assault along with rapid movements of body parts severely affects the propagation path. Wi-Dog continuously records and processes the raw CSI measurements based on a set of advanced processing algorithms. The precise spectrogram of motion-induced Doppler Effect can be recovered through a band pass filter,

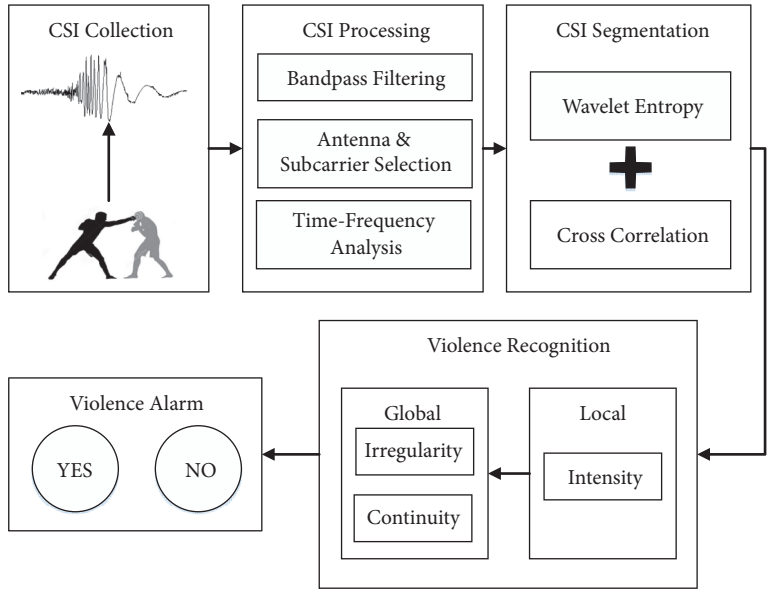


Figure 1: The architecture overview of Wi-Dog.

subcarrier selection steps, and time-frequency analysis. As a consequence, the noise-covered motion information comes out and takes up notable portion of signal fluctuations. In the CSI segmentation step, we advance the existing variance-based segmentation approach by proposing a novel assault indicator combining wavelet entropy and cross correlation, that is, a key step for Wi-Dog to detect abnormal transitions within a wide range and accomplish the goal of location-independent monitoring during human interactions. In the assault recognition step, we explicitly analyze the suspected assault series from both local and global aspects. For local analysis, we extract features representing high-intensity and adopt a SVM classifier to preserve suspected time slices. To enhance the efficiency and robustness of Wi-Dog, we take irregularity and continuity into consideration to further analyze the long-term fighting process. Wi-Dog triggers assault alarm only when suspected time slices have been confirmed by local-global analysis. Otherwise, no assault is detected within the surveillance area.

4. Wi-Dog Design

In this section, we give out critical observations of physical assault and further detail the methodologies of Wi-Dog by experimental studies.

4.1. CSI Processing

4.1.1. Band Pass Filtering. Raw CSI measurements contain high amplitude impulses, burst noises with high frequency, and significant static interferences with low frequency. To obtain sanitary CSI data with a target frequency shift, a three-order Butterworth band pass filter is a natural choice to remove irrelevant signal components in $|H(f, t)|^2$. Practically, considering real physical assaults are extremely intensive (according to [27], the average torso speed v_t of Olympic boxers can reach to $3m/s$, corresponding to $f = 2v_t/\lambda = 120\text{Hz}$), we set the lower cut-off frequency to 1Hz to eliminate interference of static components, while the upper cut-off frequency is set to 140Hz to keep more high-frequency action information.

4.1.2. Antenna and Subcarrier Selection. Observation I: Embracing multiple antennas intuitively enhances the spatial granularity by harnessing frequency diversity [28]. However, taking all subcarriers into account is not panacea for the exposure of some vital information. As is shown in Figures 2(a) and 2(b), we note that the antennas with higher variances in static are likely to possess less dynamic environment responses, which indicate that background noises contribute a lot to the captured variations of CSI measurements, rather than dynamic responses. Therefore, the variance of CSI can be helpful to exclude insensitive antennas.

Observation II: Based on the intuition that motion-induced variations of CSI waveforms are correlated, the most common variations of all subcarriers can be extracted by Principle Component Analysis (PCA). However, two challenges arise as well. First, residual noises are stubborn and nonnegligible due to their internal correlation. In Figures 2(c) and 2(d), a pair of antennas with lower cross correlation of 30 subcarriers in static is likely to possess higher correlation in dynamic environment, and vice versa. The reason is that the same subcarriers of different antennas are affected by the background noises to varying degrees, leading to similar variations in waveforms, which we term as 'relevant noise'. Second, it is a dilemma to select the first (noisy but principle information remained) [18] or the second (sanitized but with information loss) principle component [18]. To kill two birds with one stone, we meet these challenges by utilizing cross correlation to select subcarriers which are robust to relevant noises and sensitive to human movements.

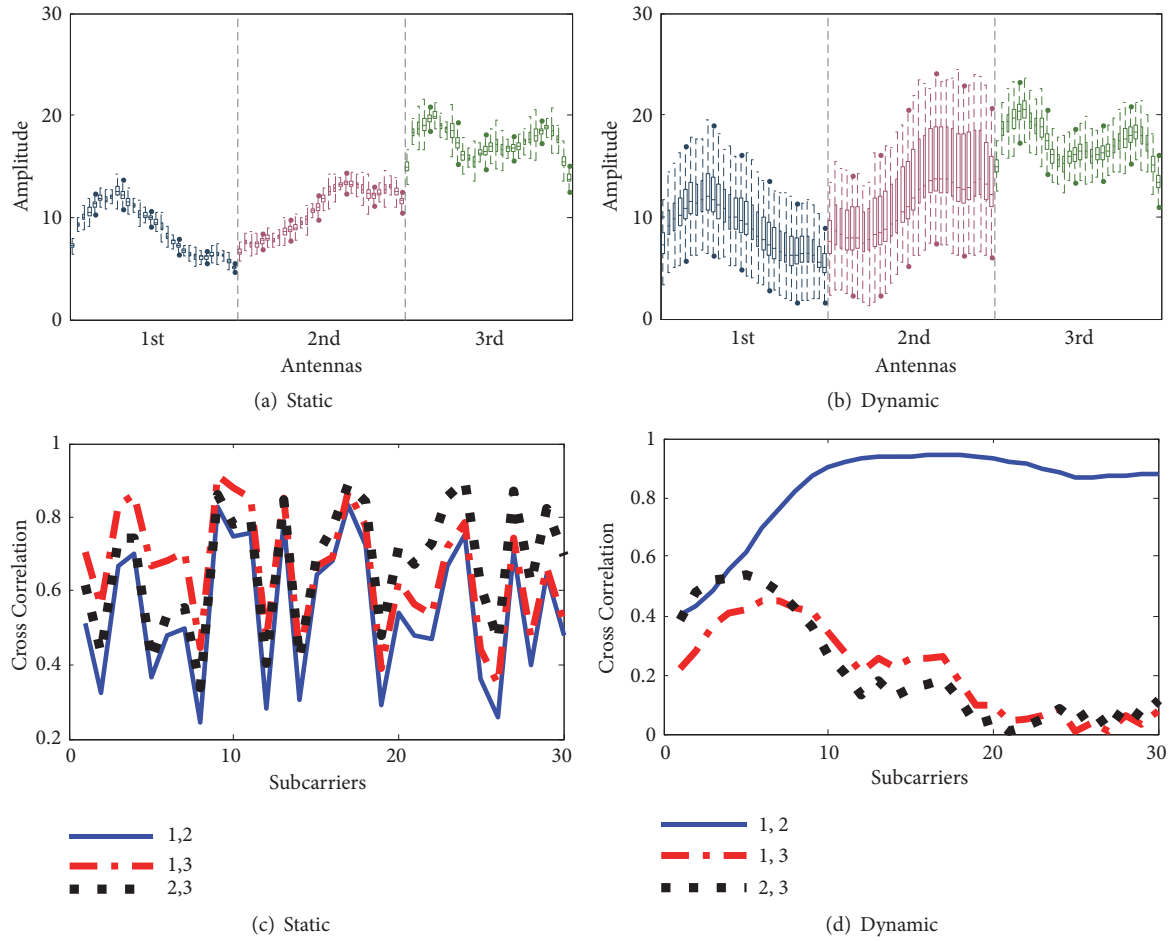


FIGURE 2: The amplitudes and cross correlations in two states.

Selection Strategy. We present our strategy for selection of antenna pairs and subcarriers as follows. For antenna selection, we first calculate the standard deviation σ and mean cross correlation C for each antenna before motion starts. Then, the Antenna Selection Indicator (ASI) is defined as $C\sigma$. The antenna pair with the lowest ASI is selected as the robust antennas. The brief break during physical conflicts, e.g., observing response, quarrelling, or any relatively slight interactions, can be termed as static moments to calibrate the antenna selection. For subcarrier selection, to find a trade-off between sensitivity and robustness, we choose the top 20 subcarriers with the lowest cross correlation values in static and with the highest cross correlation values in dynamic environment, respectively. The final subset of subcarriers are chosen from the intersection of 40 subcarriers. Specifically, the criterion of our subcarrier selection strategy is not to consider those subcarriers with the largest variance in dynamic moments, because we are unaware beforehand of whether the variation is induced by location-variant movements or relevant background interferences. What is more, to reduce dimensions and useless interferences of CSI data, we apply PCA to extract common variations of selected subcarriers and choose the first principal component for time-frequency processing. By our advanced selection strategy, the extracted first principal component contains 95% major motion information as well as negligible noises and contributes to the increase in the proportion of reflecting power in recovered spectrogram. Figure 3 shows the PCA result of all 90 subcarriers, which still contains inferior subcarriers and nonnegligible 'relevant noise'. Clearly, our processing method provides a smoother waveform with informative fluctuations and weakly noises.

4.2. Time-Frequency Analysis. To effectively analyze the CSI waveforms in the time-frequency domain, Short-Time Fourier Transform (STFT) is adopted to generate the spectrogram which shows the energy of each frequency component with time. We use the normalized FFT magnitudes and a sliding window approach for suitable time-frequency resolution of 1.93Hz and 0.5ms. A Gaussian window with a size of 3 is further applied to smooth the spectrogram. Figure 4 shows fine-grained spectrogram with two volunteers imitating real physical collisions. In this experimental study, one volunteer who plays the role of attacker is asked to aggressively push the other volunteer who acts as a victim walking toward him, while the victim is asked to respond to the assaults as real as possible. The imitated actions repeat five times, with no interruption. Triple heuristic observations have been verified

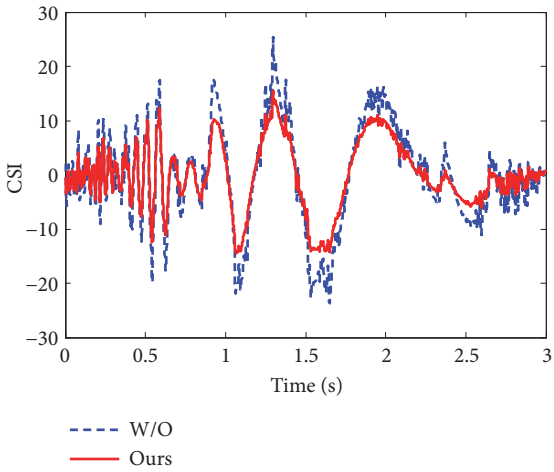


Figure 3: The PCA results with or without subcarrier selection.

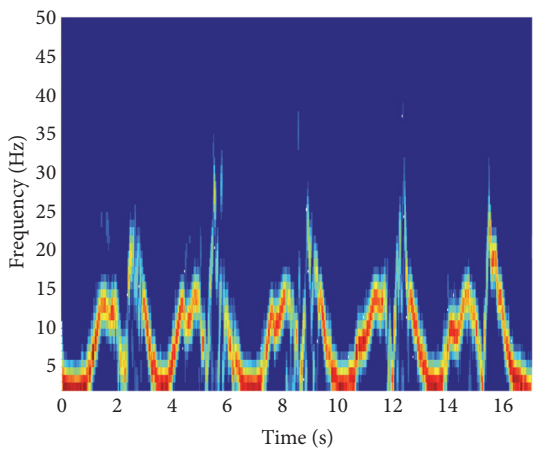


Figure 4: Constructed spectrogram with normal activities and physical assault.

based on this experimental study. (1) Drastic conflicts always contain abundant *intensity* cues (e.g., speed, acceleration, and kinetic energy) which lead to distinct variations of power in corresponding frequency bands. (2) With the escalation of assault, the attacker would bare his teeth and claws, while the victim may fight back or flee; the distribution of energy of reflected signal is *irregular* due to complex and rapid movements of body parts. (3) The physical assault could burst out anytime with a short duration, while the assault process keeps *continuity* until the attacker radically releases his resentment. These observations can be summarized as three critical characteristics, that is, intensity, irregularity, and continuity. Hence, we are inspired to exploit these unique characteristics of physical assault to realize passive assault monitoring.

4.3. CSI Segmentation. To identify the assault-induced variations, the key issue is to discern every transition points in CSI time series. Previous variance-based sliding window approach is widely adopted for passive motion detection [21]. We discard this approach in assault monitoring for two reasons. First, over-threshold variance is easily affected by irrelevant motion information, which may delay or mislead the real-time detection in target frequency. Second, the absolute value of variance cannot precisely reveal the intensity and complexity of interactions. Cross correlations between different subcarriers have been shown to be effective in reflecting the process of human walking. However, we argue that it is insufficient to discern fluctuation caused by micro-energy response from macro-movements in the distance.

Therefore, we resort to wavelet entropy (WE), a new method of complexity measurement for signals [29, 30], which is capable of monitoring micro-energy response and quantifying the order-disorder states of the reflected signals. Lower WE denotes the simpler components of frequency and more orderly changes of movements, and vice versa. WE possesses unique advantages as follows. On the one hand, compared with the periodic motions of human objects, the signals reflected from drastic actions of body parts are highly nonstationary and correspond to more randomness in time and frequency domain. WE could precisely reflect the intensity level of energy variations of target frequency bands. On the other hand, WE is a feasible and location-independent indicator, which is suitable for our dynamic experimental scenes. Wavelet entropy is defined as $WE = -\sum_{j<0} p_j \ln(p_j)$, where p_j represents the normalized ratio of wavelet energy E_j at the j-th scale, $\sum_{j=-1}^{-N} p_j = 1$.

In practice, we select db6 wavelet due to its better orthogonality and compact support, which makes wavelet transform more suitable for signal oddity detection [31]. We adopt a sliding window method to calculate WE with the sliding window width of 1s, half of which is used as the step size. To further improve the robustness of detection system, we propose a light-weight Assault Detection Indicator (VDI) combining wavelet entropy WE and cross correlation C as $VDI = (\max(WE) - WE)e^C$, with a 5-point median filter to smooth the curve. The higher VDI means the higher intensity level of activities and the more frequent signal changes. We select transition points which are the local minimum values below the predefined threshold which will be further evaluated in evaluation part.

As is shown in Figure 5, we compare the performance of our proposed method with RT-Fall's segmentation step, which is capable of real-time and continuous fall detection by variance-based segmentation. The entire procedure of human interactions lasts about 50s that consist of assault part (6s~ 30s) and normal part (31s~ 48s). The upper figure shows the normalized sliding curve combining average value and variance of amplitude. We notice that the previous segmentation method cannot precisely and timely reveal the high-intensity with the escalation of physical assault, while amicable interactions near the links may generate great variations in waveforms (37s- 40s), which is unreliable for passive assault monitoring. The lower figure reveals the superiority of our advanced segmentation approach, which clearly depicts the complexity level of assault-induced signals and mitigates the location-variant interferences caused by human normal activities.

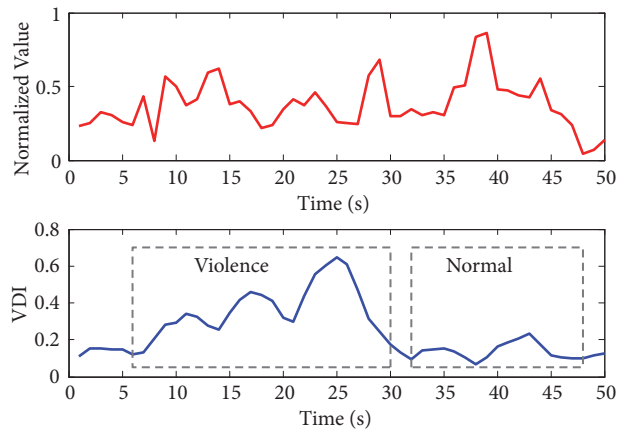


FIGURE 5: Constructed spectrogram with normal activities and physical assault.

4.4. Assault Recognition. The assault recognition step aims to explicitly characterize the features of violent attacks and precisely detect the presence of violent events. Hence, we identify suspected violent events from two perspectives, i.e., the *local view* and the *global view*. The purpose of local analysis is to identify drastic actions from mild human activities, whereas global analysis is to evaluate the irregularity as well as continuity of signals in a given time interval.

4.5. Local Analysis. The most essential part of local analysis is to obtain high-intensity information. Noting that dominant power strengths of frequency bins caused by human torso reflect the major trends of human movements, we adopt the percentile method, an effective PLCR extraction method, to estimate the torso movement speed [32]. The cumulated percentage $P(f, t)$ of energy $F(f, t)$ at a given frequency f and time t is calculated as

$$P(f,t) = \frac{\sum_0^f F(f,t)}{\sum_0^{f_{\max}} F(f,t)} \tag{5}$$

where selected frequency values f should not be singular at time t and satisfy $P(f,t) \geq 0.75$. Therefore, several features revealing high-intensity can be extracted from both the total spectrogram and dominant speed.

(1) *The area of the surrounded curve (ASR)* denotes the area of the speed curve surrounded with horizontal axis. The rationale is that assault-like actions produce relatively enormous velocity along with high peaks of amplitudes in a short time duration. The efficiency of ASR depends on previous segmentation step.

(2) *The power changing rate (PCR)* is proposed to reflect the increase of kinetic energy based on the observation that abnormal drastic motions typically have high Doppler energy content within a specific frequency band. We give the mathematical formula of PCR as follows:

$$PCR = \frac{\left|\sum_{t=t_0-1}^{t_0} \sum_{f=f_l}^{f_u} F(f,t).f - \sum_{t=t_0}^{t_0+1} \sum_{f=f_l}^{f_u} F(f,t).f\right|}{\min\left(\sum_{t=t_0-1}^{t_0} \sum_{f=f_l}^{f_u} F(f,t).f, \sum_{t=t_0}^{t_0+1} \sum_{f=f_l}^{f_u} F(f,t).f\right)} \tag{6}$$

where $F(f,t)$ represents the FFT power coefficients of a specific frequency f at time t, f_u and f_l is the upper bound and lower bound of the interested frequency band, and t_0 refers to the time when transition point is detected.

(3) *Peak amplitude bandwidth (PAB)* chooses the 1/2 and 1/4 peak amplitude bandwidth to reflect the divergence between peak values and valley values of FFT magnitudes. The reason is that energy of intensive assaults disperses in a wide band around the frequency of peak amplitude.

(4) *High-frequency duty ratio (HDR)* is often used to measure the ratio of high level of the time. Considering violent actions are always along with rapid high-frequency changes, we count the number of times when FFT coefficients $F(f,t)$ meanwhile exceed preset frequency f and a predefined threshold.

The SVM classifier is then applied to select assault-like actions, which is originally designed for binary classification. We use LibSVM toolbox [33] with Gaussian Radial Basis Function (RBF) kernel in the training process and set the cost parameter C and gamma g in kernel function to be 4 and 0.0884 through 10-fold cross-validation.

4.6. Global Analysis. Global analysis starts to be considered only when assault-like action is detected. We adopt two features to characterize the continuity and irregularity of complex physical assault.

(1) *Detection Confidence* is calculated to reflect the *continuity* of human activities. As is shown in Figure 4, violent events always last for a relatively long time due to the escalation of physical conflicts, while other normal movements (e.g., lying down and sitting down), even with abrupt acceleration, seem unlikely to occur several times within a short time duration. To quantify the continuity of actions, let J be a sequence of segmented slices; the predicted assault probabilities of following segments are calculated. We give a detection confidence C_T as

$$C_T = \text{sgn}(\text{T}) - \prod_{j \in J} \left(1 - P_j\right) \tag{7}$$

where P_j denotes the probability of violent action in the j-th segments. The formula indicates that with the increase of the number of assault-like segments within time duration T, the possibility of assault events will be higher, which could efficiently reduce the rate of false alarm. We set time duration T as 15s for real-time and robust monitoring.

(2) *Lempel-Ziv complexity* is a feasible indicator to reflect the irregular degree between the vicinity segments. Physical assault may result in irregular patterns due to various extents of attacks, while rapid macro-movements, e.g., running and frog leaping, have repetitive profiles. In such case, Lempel-Ziv complexity [34] is appropriate to measure the dissimilarity of both sanitized amplitudes and extracted velocity, by counting the number of distinct patterns and recurrence rate in target series. The normalized Lempel-Ziv complexity $C(n)$ can be calculated as

$$C(n) = \frac{c(n)}{n} \log_{\partial} n \tag{8}$$

where $c(n)$ denotes a complexity counter; ∂=2 means the binary consideration here. Finally, another SVM classifier is adopted to differentiate real physical assault from similar dynamic actions, using the normalized feature to quantify the continuity and regularity. Wi-Dog would trigger the alarm only when both preset conditions meet the criteria.

5. Evaluation

In this section, we interpret the experimental strategies, overall performance and detailed impact study of Wi-Dog, respectively.

5.1. Experimental Strategy. Experimental setup. We use a ThinkPad X200 laptop with a single antenna as the sender and a Lenovo T460 laptop with three antennas as the receiver. Both laptops are slightly modified with Intel 5300 NICs and set up to inject in monitor mode on channel 165 at 5.825 GHz, with package rate set to 1000 pkts/s. In Figure 6, we evaluate Wi-Dog with 10 volunteers (5 males and 5 females) in a multipath-affected classroom surrounded with tables and chairs, and a narrow corridor with the width of 1.5m, which can be regarded as LOS environment. We place the sender and the receiver at the height of 0.8m corresponding to the height of human torso and separate them by a distance of 3m for an appropriate detection range.

Data collection. We collect abundant CSI data in almost two weeks for (1) 5 mild interactions (i.e., shaking hands, making bows, talking with body language, hugging, and giving high-five), (2) 5 normal human actions (i.e., lying down, sitting down, walking, running, and playing exergames) performed by one volunteer with the other one standing by, and (3) long-time physical assaults imitated by two volunteers of each group with necessary protective equipment (e.g., pushing over the victim). Each volunteer is asked to continuously finish specific normal actions and then conduct assault attack in total 5 minutes for 50 sets, while there are no predefined constraints for physical assaults. The ground truth is acquired according to the video recordings. Due to the layout limitation, Figure 7 only shows an example of intensive action. At the beginning, the attacker and the victim both keep static status in Figure 7(a). Then the attacker walks toward the victim (Figure 7(b)) and pushes over the victim on the ground (Figure 7(c)). The system would be awaken and keeps a close watch on the subsequent actions. If the victim struggled to strike back and the conflict arose, Wi-Dog would consider these actions as physical assault and raise the alarm.

Metric. We evaluate the performance of Wi-Dg based on two metrics.

(i) *True Detection Rate (TDR)* is the probability that the physical assault is accurately detected from similar activities, which is defined as the ratio of accurately detected assaults from all human interactions. Higher TDR represents the effectiveness of an emergency alert system.

(ii) *False Alarm Rate (FAR)* is the proportion of the system wrong alarms when there is no violent assault happening, which can be defined as the ratio of wrongly detected assaults and accurately detected assaults. Lower FAR promises fewer misalarms and optimizes the public resource.

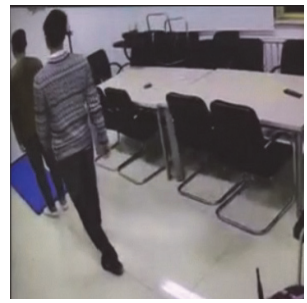
5.2. Overall Performance. In this part, we resort to three state-of-the-art anomaly detection methods, RT-Fall [21], WiFall-2014 [35], and WiFall-2017 [36] as the baselines. WiFall-2014 [35] was the first work for passive fall detection with COTS Wi-Fi devices, and its extended version WiFall-2017 [36] advanced the performance by considering the subcarrier sensitivity and principal component extraction. By comparison, we could understand the necessity of subcarrier selection. RT-Fall is the most similar work, which extracted features from bimodal CSI information and realized real-time segmentation. However, even though physical assaults and fall-like actions happen with similar energy variations, these fall detection methods cannot be directly applied due to the lack of assault-procedure analysis. We then add global analysis to further process extracted features. The main purpose of comparison is to reveal the superiority of our CSI processing algorithms and segmentation methods. As is shown in Figures 8(a) and 8(b), we notice that WiFall-2014, without subcarrier selection and action segmentation, cannot realize robust assault monitoring with only statistical features, which could mislead the assault alarm. The average TDR and FAR of WiFall-2014 in LOS are around 0.71 and 0.28, while in NLOS environment they are around 0.64 and 0.35. WiFall-2017 emphasized the importance of subcarrier selection and earned a 0.19 (0.13) higher TDR and 0.07 (0.10) lower FAR compared with its previous version in LOS (NLOS). Yet the lack of segmentation still limits its practical use. RT-Fall reveals its satisfactory performance in complex scenarios. The TDR and FAR of RT-Fall are 0.91 and 0.13 in LOS, 0.86 and 0.15 in NLOS. Compared with RT-Fall, Wi-Dog outperforms RT-Fall by 3% higher TDR and 5% lower FAR in LOS, similar TDR and slightly lower FAR around 0.85 and 0.11 in NLOS. We explain why amplitude-based Wi-Dog could achieve similar or even better performance than full-information-employed RT-Fall. First, Wi-Dog uses the selected cross correlation to characterize the properties of phase difference which has been proved in [14]. Second, Wi-Dog overcomes the drawback of amplitude-based detection method which is, as commonly argued, its vulnerability to NLOS environments based on the appropriate subcarrier

(a) Corridor

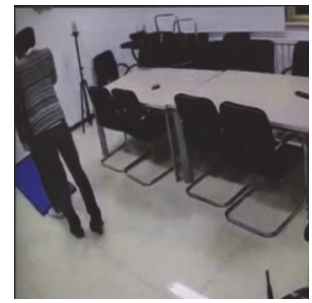
(b) Classroom

Figure 6: Experimental scenarios.

(a) Static status



(b) Walking toward the victim



(c) Pushing over

Figure 7: Video snapshots of intensive action.

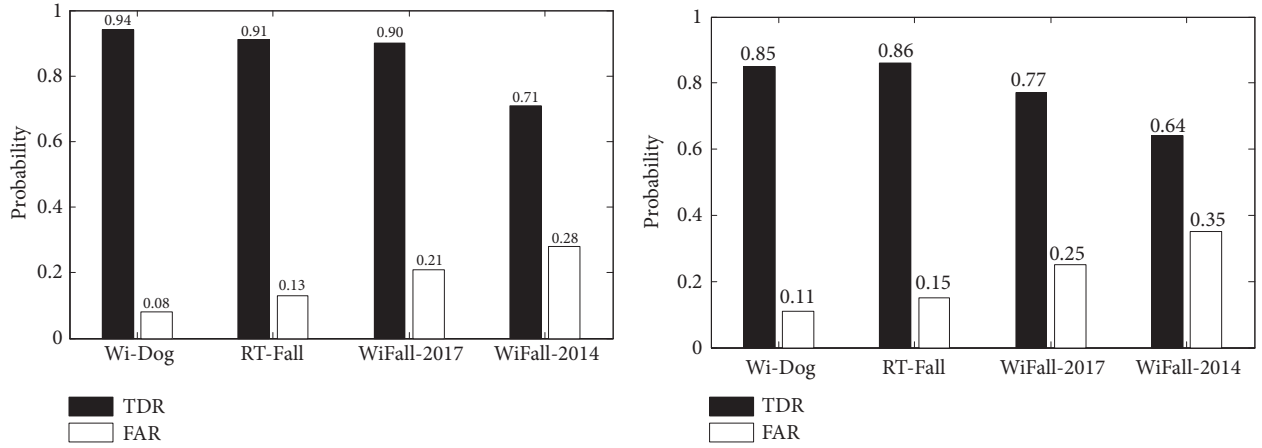
selection. Third, Wi-Dog is robust to micro-movements near the links and sensitive to macro-movements in the distance by monitoring the complexity level of suspected actions in target frequency bands. In contrast, location-variant movements may constrain the accuracy and robustness of RT-Fall.

5.3. Parameter Study. Impact of local-global analysis. In order to validate the practicability and necessity of the detailed description of assault characteristics, we evaluate the local-global analysis by separately evaluating all three properties. Table 1 shows the specific value of independent-processed part. For single-variable analysis, both intensity (*In*) and irregularity (*Ir*) analysis are infeasible to sensitively detect or precisely alarm, while only considering continuity (*Co*) is meaningless. For double-variable analysis, we notice that the method of (*In+Ir*) could achieve obvious progress in sensitivity, while it still maintains relatively high FAR in experiments. (*In+Co*) sharply lowers the FAR by excluding those slices with enormous short-time energy, while leading to the omission of assault alarm. We also take (*Ir+Co*) into consideration; while the irregular features of short-time slices are subjected to continuity analysis, the final result is not unsatisfactory. To make overall quantification of physical assault, an integrated local-global analysis is imperative for raising the sensitivity as well as reducing the false alarms.

Impact of group diversity. To evaluate the general applicability of Wi-Dog for different users, we recruit 10 volunteers consisting of 5 males and 5 females to show the impact of group diversity. During the experiments, the participants possessing various heights, weights, and ages are required to be paired with different partners. The reason is that different participants may react with various intensive extents. As is shown in Figure 9, Wi-Dog monitors physical assault of all participants with relatively high accuracy. Among the results, the detection rates of Male-beat-Male (M⟶M) and Male-beat-Female (M⟶F) seem satisfactory. We owe it to the rapid macro-movements caused by men-characters. Conversely, we also notice that in the process conducted by female, the detection rate drastically decreases to 0.88 in LOS and 0.86 in NLOS when the victim is female, and 0.85 in LOS and 0.78 in NLOS when the victim is male, which are still practicable even in real environment. We explain that Wi-Dog is sensitive to high-intensity actions and body profiles in the monitoring area, while female participants fail to maintain consistent high-intensity attacks to overweight

Table 1: Performance of local-global analysis.

Method	LOS		NLOS	
	TDR	FAR	TDR	FAR
In	0.61	0.35	0.58	0.41
Ir	0.58	0.44	0.52	0.46
In+Ir	0.81	0.23	0.75	0.28
In+Co	0.61	0.13	0.58	0.15
Ir+Co	0.73	0.25	0.66	0.35
In+Ir+Co	*0.94*	*0.08*	*0.85*	*0.11*



(a) LOS (b) NLOS

Figure 8: Performance comparison in (a) LOS and (b) NLOS.

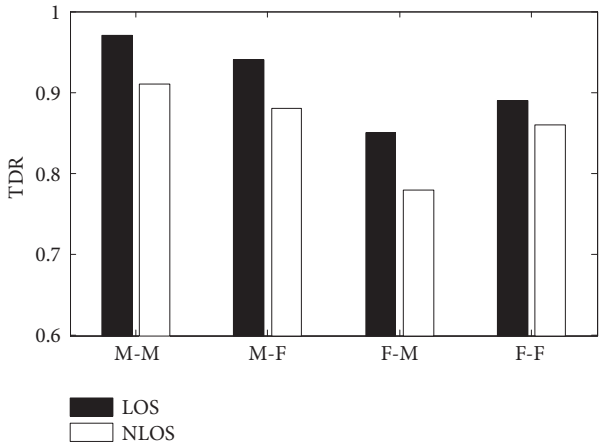


Figure 9: The impact of individual diversity.

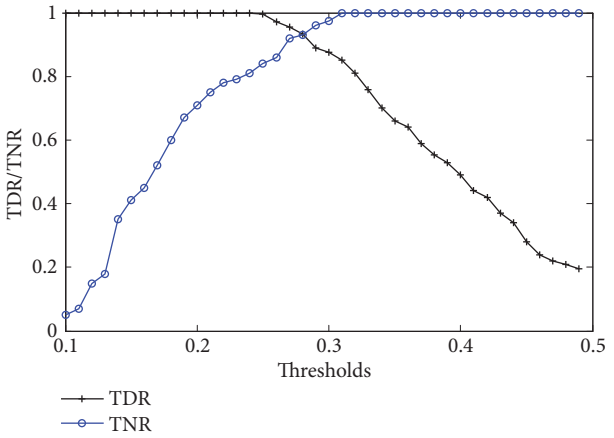


Figure 10: The impact of VDI thresholds.

male victim. The loss of fighting information may induce a slighter Doppler frequency shifts that can be confused with those normal interactions.

Impact of thresholds. As precise segmentation is essential for long-time assault detection, an appropriate threshold of VDI is needed to sensitively monitor abnormal transitions. Figure 10 depicts the TDRs and True Negative Rates (TNRs) of assault activities under a rational variation of VDI thresholds with the increment value of 0.01. With the increase of thresholds, the TDR decreases from 1 to 0.2, indicating that only extremely strenuous actions over the predefined thresholds can be correctly identified, while those below-threshold violent actions would lead to a general uptrend of TNRs. We note that lower thresholds lead to higher sensitivity but unreliability in assault monitoring. For balanced overall performance, Wi-Dog achieves the best trade-off between TDR and TNR of 0.94 and 0.92 by setting the threshold of VDI as 0.275, which is a general threshold fitting the majority of assault types.

Impact of duration time. We further evaluate the performance with changing duration time T. Intuitively, longer duration time could contain more underlying assault actions, which would significantly reduce the misinformation and ensure the system accuracy. As is shown in Figure 11, we observe an ideal trend of rising TDR as well as decreasing FAR with longer T. The reason behind observations is that assault-like actions (e.g., fall and run) bursting out enormous short-time energy can be excluded by continuity analysis, which may otherwise induce severe misalarms. However, when duration time T exceeds 20s, FAR has an obvious rising trend while TDR has a slight decline. We explain that even physical collisions would generate some similar patterns and present some kind of regularity. Furthermore, longer time duration would bring about relatively higher possibility of confusion with continuous middle-intensity movements during exergames. We set T=20 as a reasonable choice by fully considering the real-time capability and accuracy.

Impact of packet rates. Based on a clean channel 165 and both laptops set up in monitor mode to avoid uncontrolled packet losses, further experiments are conducted to see the relationship between system accuracy and sampling rate. Since abundant information of drastic actions can be fully

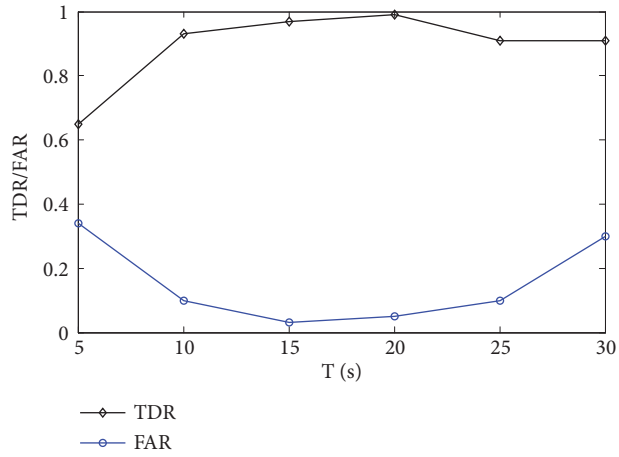


Figure 11: The impact of duration time.

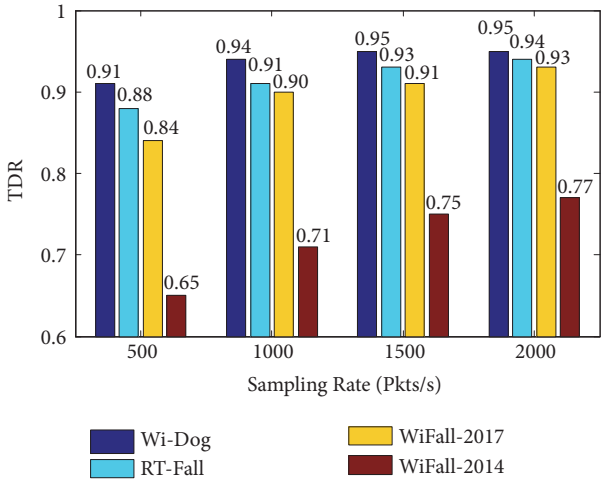


Figure 12: The impact of packet rates.

preserved by fine-grained transmission, we suppose that Wi-Dog can achieve better performance with the increase of sample rates. Figure 12 verifies our intuition, which shows the positive correlation between two variables. By horizontal comparison, we notice that all systems enhance the performance in varying degrees. The rationale is that growing packets in fixed time window would induce more informative features, e.g., wavelet entropy and variances. By vertical comparison, we observe that Wi-Dog outperforms the other two methods when the sampling rate increases from 500 pkts/s to 2000 pkts/s. In particular, Wi-Dog possesses a favourable result of 0.91 even with only 500 pkts/s, while the other three methods raise the TDR but could only approximate the value in 2000 pkts/s. Therefore, Wi-Dog could make assault monitoring available with imperfect Wi-Fi devices even with low sampling rate.

Impact of distance to LOS path. In general, amplitude-based detection severely suffers from its natural deficiencies, for example, obstacle and long-distance attenuation, which limits the effective detection range. We test the impacts of distance with two baselines in the mid-perpendicular of a single link, ranging from 1m to 5m. Figure 13(a) demonstrates the varied trend with the changing performing distance. Obviously, the detection rate of WiFall-2014 suffers from maximum long-distance attenuation due to inferior subcarriers. With efficient signal processing steps, the other three methods keep relatively high detection rate. Due to the lack of accurate segmentation, WiFall-2017 omits some useful segments and only obtains 0.74 at 5 meters; RT-Fall and Wi-Dog are both competitive to sensitively detect physical assault.

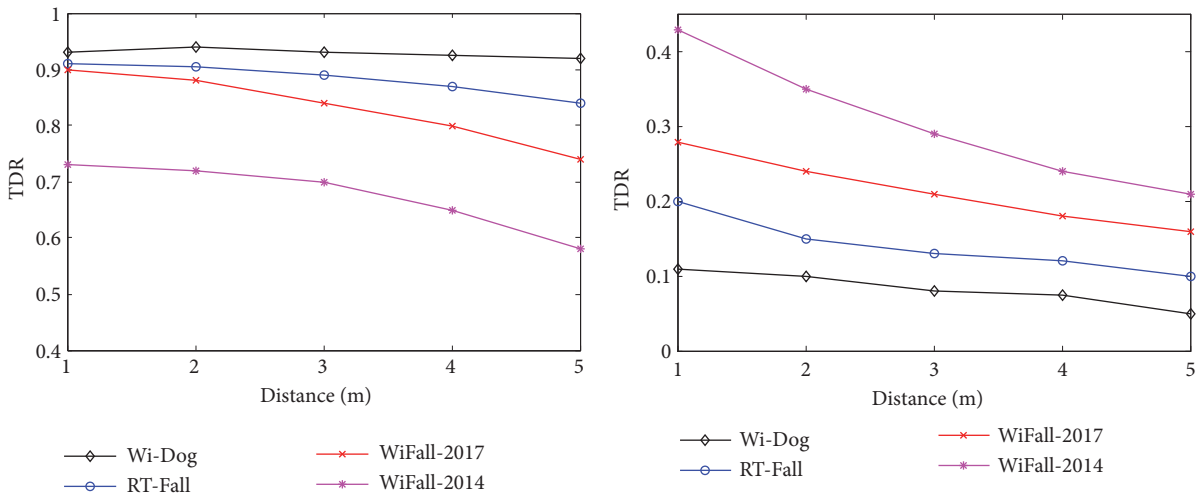
In Figure 13(b), we notice that the closer the distance between users and links is, the higher the FAR is. This is because the location of users is relevant to the sensitivity of the monitoring system. For example, some slight interactions near the link could be easily mistaken as intensive actions, so the system may raise misalarm. The reason lies in the variance-based segmentation, which is a location-dependent indicator. In contrast, Wi-Dog copes well with this critical challenge by decomposing superposed signal into different frequency band, which makes Wi-Dog practicable in wide range monitoring.

6. Discussion and Limitations

6.1. Detecting Multiple Targets in Close Contact. Given that Wi-Dog excels in monitoring physical assault mainly based on its natural properties of high-intensity, irregularity, and continuity, we intend to further promote the sensitivity of Wi-Dog by detecting multiple targets in close contact. The original motivation springs from vision-based assault method [37] by effective spatiotemporal modeling to detect crowd abnormal events, which implies tight connection between crowd density and probability of abnormal events. Unfortunately, two thorny problems arise.

First, as the latest work [38] counts crowd by using multiple wireless links, it seems unsolvable to calculate the crowd density in a single link beforehand; let alone the relative locations. TensorBeat [39] is recently proposed to monitor multiperson breath rates in the high dimensions. Wi-Run [13] borrows the idea of tensor decomposition for multiple-runner recognition. However, the calculation of rank R representing the number of people is an NP-hard problem in tensor decomposition. Second, the obstacle caused by human motions in LOS path cannot be captured due to our reflection-based theory basis, though the problem can be mitigated by optimal device placements in experiments. Nevertheless, we leave the early detection of multiperson scenario as one of our future works.

6.2. Detecting Multitype Assault. Although Wi-Dog has been validated to accurately monitor physical assault based on the rapid and continuous movements of body parts, it still cannot cover all types of violent actions in real world. For example, if victim encounters sudden deadly gunshots or any single lethal assault, the assault alarm would not be triggered. Note that assault process comprises various features, including audio cues and visual cues [40] (e.g., screaming, gunshots, explosion, and blood); just in case, multiple sensors should be put into services and provide



(a) TDR comparison

(b) FAR comparison

Figure 13: The impact of performing distance.

multimodal information for major-concerned public places. We envision an enhanced version of Wi-Dog by multimodal feature processing for more complicated scenarios, in which wireless sensing technology acts as an avant-courier and decision-making center.

6.3. Embracing Deep Learning Methods. The feasibility of Wi-Dog partly relies on the elaborate-design spectrum features. Specifically, we extract the spectrum features to represent the motion intensity and further consider the irregularity and continuity so as to depict the correlation of motions, while it requires the domain knowledge and handcraft feature selection. Moreover, if two users push over along the tangent line of the Fresnel Zone [17], the reflected signals could be very weak and the spectrum features could be obscure. Therefore, TDR is relatively low (i.e., at almost 0.65). Global analysis is helpful to some extent since the directions of human movements are irregular, yet it lowers the time-efficiency. It is also a natural choice to deploy multiple pairs of trans-receivers [21, 35, 36], yet it increases the economic cost.

Recent years show the potential of deep learning embedded Wi-Fi sensing, such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN), which could automatically extract high-level prominent features through multiple hidden layers. We believe the Wi-Dog could be further promoted by using these advanced models. For example, intrinsic motion characteristics could be captured by CNN through convolutional operations, and the long-term relations could be established by RNN. Since there is no work focusing on deep learning-based abnormal detection, we leave it as our future work.

7. Related Work

Wi-Dog is related to the previous abnormal detection works in three categories: wearable sensor based, camera based, and wireless sensing based.

Wearable sensor based. Wearable sensor-based systems are both widely used and commercially available owing to the rapid development of sensor technology.

For example, [4] recognized free-weight activities by attached RFID tags and further assessed the exercise quality from a local-global perspective. Reference [2] developed a fall detection system by monitoring the variation of 3-dimension acceleration data, with a specific wearable device placed on human's waist. Reference [3] utilized a multisensor integrated glove to identify abnormal behaviours of paralysis patients. Reference [5] explored the possibility of using a smartphone to detect abrupt physical attacks. However, all these methods require per-user wearing sensors, while Wi-Dog aims to achieve contact-free assault detection.

Camera based. Camera-based assault detection system uses fixed cameras to capture pictures or video frames to identify human assault. Reference [6] firstly presented an approach to analyze assault relying on the information of motion trajectory and acceleration. Reference [7] further studied aggressive fighting using extreme acceleration pattern as discriminant feature. Reference [8] designed a novel image descriptor for assault with spatial and temporal features. Even so camera-based schemes address the problem of wearing extra devices, the issues of privacy and limited monitoring scope are still up in the air.

Wireless sensing based. Wireless sensing-based schemes attract extensive attention in recent years [9–11]. We adopt CSI-based schemes because fine-grained CSI is available in ubiquitous Wi-Fi devices. In recent studies, CSI was utilized to track the walking path [41] as well as functional body movements [12] and recognize walking postures [14], multiperson running detection [13], and even slight movements like breath [17] and finger movements [18–20]. However, existing CSI-based schemes highly depend on the observation of the repetitions with reproducible features, while violent behaviours can be random and irregular. NotiFi [42] proposed a non-parameter training scheme for abnormal activity detection by using Dirichlet process. Yet it

requires the user's continuous walking and cannot handle the presence of multiple users. RT-Fall [21] is the most similar state-of-the-art, which used the variations of calibrated CSI phase differences to detect fall and extracted the distinct power decline pattern to find transition points. However, we discard the phase information of CSI power due to the impact of carrier frequency offsets. Motivated by CARM [15] which extracted PLCRs from time-frequency analysis, Wi-Dog makes one step further in accurate feature extraction of velocity, including a selection strategy of antennas and subcarriers as well as a segmentation method. Furthermore, Wi-Dog improves the robustness through all-round local-global analysis.

8. Conclusion

Wi-Dog is a noninvasive physical assault monitoring scheme on a single link with commercial Wi-Fi devices, which consistently analyzes the local-global characteristics of CSI waveforms. A set of novel CSI processing methods are proposed to choose reliable antenna pairs and sensitive subcarriers. Moreover, a feasible Assault Detection Indicator (VDI) is developed to monitor target frequency transitions of location-variant behaviours. Finally, all-round local-global analysis is adopted to fully exploit the features of physical assault. We prototype Wi-Dog on commodity Wi-Fi devices and evaluate the overall performance in both LOS and NLOS scenarios. Experimental results further validate the accuracy and robustness of Wi-Dog, compared with the state-of-the-art abnormal detection methods. We consider Wi-Dog as an early step toward general emergency detection on wireless sensing and a significant complement for computer-vision-based abnormal detection in security-minded places, including, but not limited to, terrorist threat warning, fall detection for the elderly, and exercise quality assessment.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 6 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by National Key Research and Development Project of China No. 2017YFC0704100 (entitled New Generation Intelligent Building Platform Techniques).

References

[1] C. B. Evans, M. W. Fraser, and K. L. Cotter, "The effectiveness of school-based bullying prevention programs: A systematic review," *Aggression and Violent Behavior*, vol. 19, no. 5, pp. 532–544, 2014.

[2] F. Wu, H. Zhao, Y. Zhao, and H. Zhong, "Development of a wearable-sensor-based fall detection system," *International Journal of Telemedicine and Applications*, vol. 2015, Article ID 576364, p. 11, 2015.

[3] A. Nelson, J. Schmandt, P. Shyamkumar et al., "Wearable multisensory gesture recognition for paralysis patients," in *Proceedings of the SENSORS 2013*, pp. 1–4, IEEE, 2013.

[4] H. Ding, L. Shangguan, Z. Yang et al., "FEMO: a platform for free-weight exercise monitoring with RFIDs," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, SenSys 2015*, pp. 141–154, ACM, Republic of Korea, November 2015.

[5] Z. Sun, S. Tang, H. Huang et al., "iprotect: detecting physical assault using smartphone," in *Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications*, pp. 477–486, Springer, 2015.

[6] A. Datta, M. Shah, and N. D. V. Lobo, "Person-on-person assault detection in video data," in *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 1, pp. 433–438, IEEE, 2002.

[7] O. Deniz, I. Serrano, G. Bueno et al., "Fast assault detection in video," in *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 2, pp. 478–485, IEEE, 2014.

[8] T. Zhang, W. Jia, B. Yang et al., "Mowld: a robust motion image descriptor for assault detection," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 1419–1438, 2017.

[9] T. Wei and X. Zhang, "MTrack: high-precision passive tracking using millimeter wave radios," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom 2015*, pp. 117–129, ACM, France, September 2015.

[10] F. Adib, Z. Kabelac, D. Katabi et al., "3d tracking via body radio reflections," *NSDI*, vol. 14, pp. 317–329, 2014.

[11] Y. Tian, G.-H. Lee, H. He, C.-Y. Hsu, and D. Katabi, "RF-based fall monitoring using convolutional neural networks," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, article no. 137, pp. 1–24, 2018.

[12] H. Zhang, J. Smeddinck, R. Malaka et al., "Wireless non-invasive motion tracking of functional behaviour," *Pervasive and Mobile Computing*, 2019.

[13] L. Zhang, M. Liu, L. Gong et al., "Wi-Run: multi-runner step estimation using commodity Wi-Fi," in *Proceedings of the 15th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON 2018*, IEEE, Hong Kong, June 2018.

[14] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 363–373, ACM, 2016.

[15] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[16] K. Qian, C. Wu, Z. Yang et al., "Decimeter level passive tracking with WiFi," in *Proceedings of the 3rd Workshop on Hot Topics in Wireless*, pp. 44–48, ACM, 2016.

[17] H. Wang, D. Zhang, J. Ma et al., "Human respiration detection with commodity WiFi devices: Do user location and body

orientation matter?" in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2016*, pp. 25–36, ACM, Germany, September 2016.

[18] K. Ali, A. X. Liu, W. Wang et al., "Recognizing keystrokes using WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1175–1190, 2017.

[19] Y. Ma, G. Zhou, S. Wang et al., "Signfi: sign language recognition using WiFi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, article no. 23, 2018.

[20] Q. Zhou, J. Xing, W. Chen et al., "From signal to image: enabling fine-grained gesture recognition with commercial Wi-Fi devices," *Sensors*, vol. 18, no. 9, article 3142, 2018.

[21] H. Wang, D. Zhang, Y. Wang et al., "RT-Fall: a real-time and contactless fall detection system with commodity WiFi devices," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, 2016.

[22] Q. Zhou, C. Wu, J. Xing et al., "Wi-Dog: monitoring school violence with commodity WiFi devices," in *Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications*, Springer, Cham, Switzerland, 2017.

[23] E. G. Krug, J. A. Mercy, L. L. Dahlberg et al., "The world report on assault and health," *The Lancet*, vol. 360, no. 9339, pp. 1083–1088, 2002.

[24] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: real-time detection of violent crowd behavior," in *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2012*, pp. 1–6, IEEE, USA, June 2012.

[25] Q. Pu, S. Gupta, S. Gollakota et al., "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking, MobiCom 2013*, pp. 27–38, ACM, USA, October 2013.

[26] D. Halperin, W. Hu, A. Sheth et al., "Tool release: gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53-53, 2011.

[27] T. J. Walilko, D. C. Viano, and C. A. Bir, "Biomechanics of the head for olympic boxer punches to the face," *British Journal of Sports Medicine*, vol. 39, no. 10, pp. 710–719, 2005.

[28] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015.

[29] Y. Wang, X. Yu, Y. Zhang et al., "Detecting and monitoring the micro-motions of trapped people hidden by obstacles based on wavelet entropy with low centre-frequency UWB radar," *International Journal of Remote Sensing*, vol. 36, no. 5, pp. 1349–1366, 2015.

[30] Y. Wang, W. Li, J. Zhou, X. Li, and Y. Pu, "Identification of the normal and abnormal heart sounds using wavelet-time entropy features based on OMS-WPD," *Future Generation Computer Systems*, vol. 37, pp. 488–495, 2014.

[31] Z. Li, J. Shen, P. Wei et al., "Voltage fluctuation and flicker monitoring system using LabVIEW and Wavelet Transform," *Journal of Computers*, vol. 5, no. 3, pp. 417–424, 2010.

[32] P. Van Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar & Navigation*, vol. 2, no. 2, pp. 135–145, 2008.

[33] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[34] M. Aboy, R. Hornero, D. Abásolo, and D. Álvarez, "Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2282–2288, 2006.

[35] C. Han, K. Wu, Y. Wang, and L. M. Ni, "WiFall: device-free fall detection by wireless networks," in *Proceedings of the IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 271–279, IEEE, Canada, May 2014.

[36] Y. Wang, K. Wu, and L. M. Ni, "WiFall: device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2017.

[37] J. Wang and Z. Xu, "Spatio-temporal texture modelling for real-time crowd anomaly detection," *Computer Vision and Image Understanding*, vol. 144, pp. 177–187, 2016.

[38] W. Xi, J. Zhao, X. Li et al., "Electronic frog eye: counting crowd using WiFi," in *Proceedings of the IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 361–369, IEEE, Canada, May 2014.

[39] X. Wang, C. Yang, and S. Mao, "Tensorbeat: tensor decomposition for monitoring multi-person breathing beats with commodity wifi," 2017, https://arxiv.org/abs/1702.02046.

[40] P. C. Ribeiro, R. Audigier, and Q. C. Pham, "Rimoc, a feature to discriminate unstructured motions: Application to assault detection for video-surveillance," *Computer Vision and Image Understanding*, vol. 144, pp. 121–143, 2016.

[41] K. Qian, C. Wu, Y. Zhang et al., "Widar2.0: passive human tracking with a single Wi-Fi link," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 350–361, ACM MobiSys, Munich, Germany, 2018.

[42] D. Zhu, N. Pang, G. Li et al., "NotiFi: non-invasive abnormal activity detection using fine-grained wi-fi signals," in *Proceedings of the 2017 International Joint Conference on Neural Networks*, IEEE, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 2350694, 9 pages
https://doi.org/10.1155/2019/2350694

Research Article

A Double Adaptive Approach to Tackle Malicious Users in Cognitive Radio Networks

Muhammad Sajjad Khan,[1] **Muhammad Jibran**,[1] **Insoo Koo**,[2] **Su Min Kim**,[3] **and Junsu Kim**[3]

[1]*Department of Electrical Engineering, International Islamic University Islamabad, Pakistan*
[2]*School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 680-749, Republic of Korea*
[3]*Department of Electronic Engineering, Korea Polytechnic University, 237 Sangidaehak-ro, Siheung-si, Gyeonggi-do 429-793, Republic of Korea*

Correspondence should be addressed to Junsu Kim; junsukim@kpu.ac.kr

Received 5 September 2018; Revised 1 February 2019; Accepted 14 March 2019; Published 27 March 2019

Guest Editor: Mohamad K. A. Rahim

Cognitive radio (CR) is being considered as a vital technology to provide solution to spectrum scarcity in next generation network, by efficiently utilizing the vacant spectrum of the licensed users. Cooperative spectrum sensing in cognitive radio network has a promising performance compared to the individual sensing. However, the existence of the malicious users' attack highly degrades the performance of the cognitive radio networks by sending falsified data also known as spectrum sensing data falsification (SSDF) to the fusion center. In this paper, we propose a double adaptive thresholding technique in order to differentiate legitimate users from doubtful and malicious users. Prior to the double adaptive approach, the maximal ratio combining (MRC) scheme is utilized to assign weight to each user such that the legitimate users experience higher weights than the malicious users. Double adaptive threshold is applied to give a fair chance to the doubtful users to ensure their credibility. A doubtful user that fails the double adaptive threshold test is declared as a malicious user. The results of the legitimate users are combined at the fusion center by utilizing Dempster-Shafer (DS) evidence theory. Effectiveness of the proposed scheme is proved through simulations by comparing with the existing schemes.

1. Introduction

Wireless network technologies are the most promising technologies in the twentieth century. Today, we already have over a dozen wireless technologies in use: Wi Fi, Bluetooth, Zig Bee, NFC, LTE, earlier 3G standards, satellite services, etc. Due to the proliferation of these wireless networks and the increase in the number of users the spectrum scarcity problem is raised. On the other hand, various reports have shown that the spectrum is inefficiently utilized such that the spectrum is underutilized at a fixed frequency and at a random geographical area [1]. Federal Communication Commission (FCC) states that temporal and geographical variations in the utilization of the assigned spectrum vary from 15% to 85% [2]. One promising solution to this problem is proposed by Joseph Mitola, i.e., "Cognitive Radio (CR)" [3, 4].

CR is a vital technology to improve spectrum utilization. A major challenge in CR is spectrum sensing that identifies the presence of the licensed/primary user (LU) in the network and whenever the LU is detected, secondary user (SU) needs to vacate the channel [5].

Sensing reliability of a single SU degraded by fading and hidden terminal problems. This problem is overcome by the use of cooperative spectrum sensing (CSS), which involves exchange of local sensing decision between multiple SUs. SUs send the sensing result to fusion center by utilizing either the hard decision or soft decision rules.

There are two types of CSS, one is centralized CSS and the other one is distributed CSS. In the centralized CSS, all SUs sense the environment and send their information about the presence of LU to the data fusion center (DFC) and the DFC gives the final decision about the presence of LU. In the distributed CSS in which there is no central node, every SU

senses the radio environment and different SUs share their information and make their own decision with distributed manner [1].

Various detection techniques are utilized in the literature to detect the presence of LU. Various detection techniques can be categorized as energy detector based sensing technique, waveform-based sensing technique, cyclostationarity based sensing technique, and matched-filtering technique [6]. Among the techniques, energy detector gives an effective spectrum sensing performance with low complexity. Energy detector is a noncoherent detector which detects the presence of the LU signal by measuring its energy and comparing it with a predetermined threshold. Furthermore, this technique does not require any prior information about LU and it is easy to implement and to be extended for the other spectrum sensing.

Meanwhile, CR networks (CRNs) are highly vulnerable to security threats. Security for wireless networks is an important part which ensures secure operation of the underlying network infrastructure [7]. Various attacks are studied in the literature, which highly degrades the performance of CRN. The most common attacks in CRN are primary user emulation attack (PUEA) and spectrum sensing data falsification (SSDF) attack [8]. In the PUEA, a malicious user behaves like an incumbent transmitter so as to enforce SUs to vacate the spectrum band. In the SSDF attack, the malicious users send false information about the presence or absence of LU to the fusion center. The SSDF attacks severely degrade the spectrum sensing reliability and spectrum utilization.

For secure CSS from the SSDF attacks, various schemes have been proposed. In [9], a scheme was proposed to prevent the SSDF attacks by calculating and updating the credit value of the SUs; malicious users are excluded to avoid the SSDF attacks in cooperative spectrum sensing. In [10], a cooperative scheme based on adaptive threshold is proposed which utilizes matched filter detector as a second stage detector in a confused region between signal and noise, and in the clear region between signal and noise, energy efficient energy detector is used as a first stage detector. Main problem of the conventional energy detector is its low detection performance at low signal-to-noise-ratio (SNR) region new approach has been proposed in [11] to solve the problem. In [11] a bilevel thresh holding approach for energy detection is proposed. In [12], an improved soft fusion-based algorithm was proposed. The authors made improvements in the traditional soft fusion algorithm by establishing the reputation mechanism according to the SU's past service qualities. The different SUs' reputation degrees are utilized for allocation of weights to the SUs in the fusion, and using this scheme the effect of the malicious users can be reduced. In [13–15], the authors explored Dempster-Shafer (D-S) evidence theory in CSS. It includes four consecutive procedures, which are (1) basic probability assignment (BPA) with this approach, (2) holistic credibility calculation, (3) option and amelioration for BPA, and (4) evidence combination via the D-S rule, respectively [15]. In [16], a new method was proposed, which estimated the attack strength and applies it in the k−out−N rule to obtain the optimum value of k that minimizes the Bayes risk.

On the other hand, above schemes do not consider the doubtful region where it is not clear whether a certain user is legitimate or malicious. While [10] considered the scenario with the doubtful region, matched filter detection technique is used. It is worthy of note that the matched filter detection has a complexity issue where it needs prior knowledge about LU.

In our proposed scheme, we utilize D-S evidence theory with double adaptive threshold method to differentiate among legitimate, doubtful, and malicious users. In the proposed scheme, weights are assigned to each user by utilizing maximal ratio combining (MRC). The weights are assigned to individual SUs, once weights are assigned and masses, i.e., basic probability assignment (BPA), are updated. The legitimate users have the highest weights and the malicious users have the lowest weights. The doubtful users have the weights between those of the legitimate and the malicious users. To ensure its credibility as a legitimate user a fair chance is given through the proposed adaptive algorithm. If it proves its credibility, it is declared as a legitimate user. Otherwise it is declared as a malicious user and withdrawn from the final decision. Through the simulation results, we verify that our proposed scheme is effective and efficient compared to the existing schemes.

The remaining of the paper is organized as follow: system model description is given in Section 2. Section 3 gives a detailed description of the proposed double adaptive threshold scheme and the proposed algorithm at the fusion center. In Section 4, we evaluate the performance of the proposed scheme and compare with the existing schemes. Finally, the paper is concluded in Section 5.

2. System Model

We consider a cognitive radio network that consists of secondary users (SUs), malicious users (MU), and a common receiver, which plays a role as a fusion center as shown in Figure 1. All SUs in the network perform spectrum sensing and transmit sensing result in the fusion center determining if a licensed user (LU) is present or absent in the network. All the SUs including both the legitimate and malicious users participate in cooperation to determine the status of the licensed user in the network.

We assume that each SU independently performs local sensing by using energy detector. The local sensing is a binary hypotheses testing under the absence H_0 or the presence H_1 of the LU in the network, which is given by

$$y\,(\mathrm{n}) = \begin{cases} z\,(n)\,; & H_0, \\ z\,(n) + s\,(n)\,; & H_1, \end{cases} \tag{1}$$

where z(n) denotes the additive white Gaussian noise (AWGN) and s(n) denotes the transmitted signal from the LU.

Since we consider the energy detection technique for the collection of information on the existence of the LU in the network, the test statistics is equivalent to an estimation of the received signal which is given by each SU as

$$x_{E_j} = \sum_{j=1}^{N} \left| y_j \right|^2, \tag{2}$$

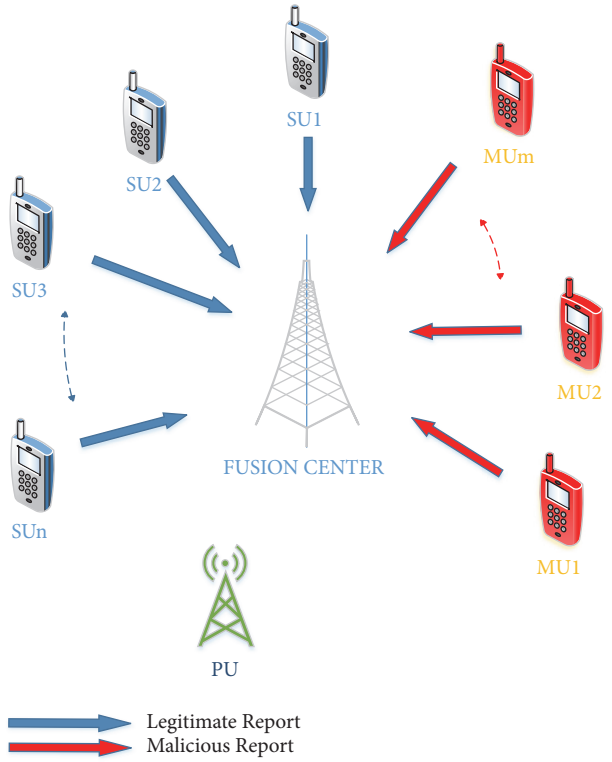


Figure 1: System model.

where $N = 2TW$, in which T is the sensing duration and W is the bandwidth, and y_j denotes the *j-th* sample of the received signal. According to the central limit theorem (CLT), when the value of N is large enough, e.g., $N > 200$, the combined signal can be well approximated as a Gaussian random variable under hypotheses H_0 and H_1, with means μ_0 and μ_1 and variances σ_0^2 and σ_1^2, which are given by [17]

$$\begin{aligned} &\mu_0 = N, \\ &\sigma_0^2 = N(\gamma + 1); \\ &\qquad\qquad H_0, \\ &\mu_1 = N, \\ &\sigma_1^2 = 2N(2\gamma + 1); \\ &\qquad\qquad H_1, \end{aligned} \tag{3}$$

where γ is the signal to noise ratio *(SNR)* the LU at the SUs.

In D-S evidence theory, the frame of discriminant A can be defined as $\{H_0, H_1, \Omega\}$, where Ω describes whether of the hypotheses is true or not. Based on parameters of means and variances, the BPA $mH_0(i)$ and $mH_1(i)$ are determined as a cumulative distribution function, respectively, by using hypotheses of the absence and the presence as follows [18]:

$$\begin{aligned} H_0 :\ \mathrm{mH}_0(\mathrm{i}) &\longleftarrow m_i\left(x_{E_i} \mid H_0\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(x_{E_i}-\mu_0)^2/\sigma_0^2}, \\ H_1 :\ \mathrm{mH}_1(\mathrm{i}) &\longleftarrow m_i\left(x_{E_i} \mid H_1\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_{E_i}-\mu_1)^2/\sigma_1^2}, \end{aligned} \tag{4}$$

where $m_i(x_{E_i} \mid H_0)$ and $m_i(x_{E_i} \mid H_1)$ denote masses of the BPA values for the absence or presence of the LU in the network.

3. Proposed Scheme

In this section, we provide detailed description of the proposed scheme. In CSS, SUs utilize an energy detection technique to sense the existence of the LU in the network. After performing the spectrum sensing, the SUs measure their mass values by using the *BPA*. Once the mass values of the SUs are measured, the SUs can be classified into three different categories. If the mass value of an SU is less than the lowest threshold (thr_1), the SU is identified as a MU and discarded from the final decision. If the mass value of an SU is greater than the highest threshold (thr_2), the SU is identified as a legitimate SU. Finally, the mass value of a SU lies between the lowest threshold (thr_1) and highest threshold (thr_2); the SU is categorized as a doubtful user. In order to provide a fair opportunity to doubtful users and ensure their credibility, we apply the proposed algorithm to those users. If the credibility is ensured, the user is considered as a legitimate SU; otherwise, it is declared as an MU and discarded from the final decision at the fusion center. The proposed double adaptive threshold is graphically described in Figure 2.

In Figure 2, it is shown a single threshold, fixed double threshold, and the proposed adaptive threshold. The single threshold does not take the doubtful users into consideration. It categorized the SU as either legitimate user or an MU, which degrades the performance of the system. In fixed double threshold, the lowest threshold and highest threshold

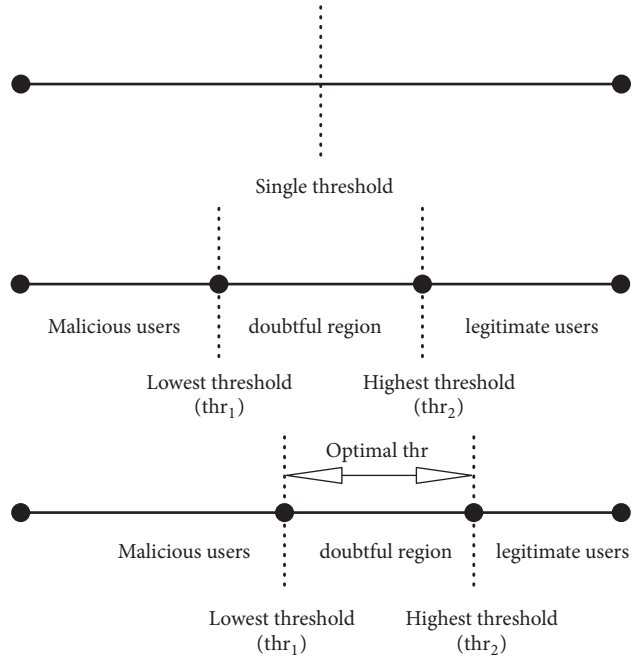


Figure 2: Proposed double adaptive threshold.

are fixed. In the proposed scheme, we consider legitimate, doubtful, and MU in a double adaptive thresholding scenario and provide a fair chance to doubtful user to prove their credibility.

The mass value of each SU is measured in (4). After the mass values of the SUs are determined, the next step is to measure the weighting factor for each SU to update the masse value. The weightage of each user is determined by

$$w(i) = \sqrt{\frac{SNR(i)}{Total\ SNR}}, \tag{5}$$

where *SNR(i)* is the signal to noise ratio of the *i-th* SU, and *Total SNR* is the sum of all the SUs' *SNRs*.

In the proposed algorithm, if the mass value of an SU is less than the lowest threshold (thr_1) then the weight assignment to the SU is zero and considered as an MU. If the mass value of an SU is greater than the lowest threshold (thr_1) but lower than the highest threshold (thr_2), its mass value is updated and compare with the proposed adaptive threshold. If its mass value is still lower than the highest threshold (thr_2), it is categorized as an MU and discarded from the final decision at the fusion center. Finally, the updated mass value is sent to fusion center for final decision.

Once the weights of each SU are measured using (5), then the mass values of SUs are updated as

$$\begin{aligned} \mathrm{m'H_0\,(i)} &= \mathrm{mH_0\,(i)} + \left(\mathrm{mH_0\,(i)\,w\,(i)}\right), \\ \mathrm{m'H_1\,(i)} &= \mathrm{mH_1\,(i)} + \left(\mathrm{mH_1\,(i)\,w\,(i)}\right), \end{aligned} \tag{6}$$

where $m'H_0(i)$ is the updated masses of hypotheses of absence of LU as reported by *i-th* secondary user and $m'H_1(i)$ is the updated masses of hypotheses of presence of LU as reported by *i-th* SU.

The performance of spectrum sensing in CR is enhanced by keeping the highest value of probability of detection (P_d) and lowest value of the probability of false alarm (P_f). According to IEEE 802.22 (WRAN), to prevent any interference between the LU and SU the probability of detection (P_d) needs to be as high as possible. To prevent underutilization of spectrum, probability of false alarm (P_f) needs to be kept as low as possible. Thus, the threshold should be selected such that we receive optimum value of P_d and P_f. Thus, by picking up different threshold values to obtain the lowest possible values P_f and highest P_d, we obtain an optimal threshold. P_f and P_d are calculates by (7) and (8), respectively:

$$P_f = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}}\left(\frac{thr\,(k) - 2m\,)}{\sqrt{4m}}\right)\right), \tag{7}$$

$$P_d = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}}\left(\frac{thr\,(k) - 2m\,(SNR\,(i) + 1)}{\sqrt{4m\,(2SNR\,(i) + 1)}}\right)\right), \tag{8}$$

where *thr(k)* is the range of thresholds.

The upper and lower limit of the double adaptive threshold is selected based on the requirement of the decision. The optimal threshold for masses is also calculated using the same formula except thr(k) is replaced by a range of masses for minimal P_f and maximal P_d.

Then, the fusion center is applied in Algorithm 1.

Once the mass values of the SUs for both hypotheses $mH_0(i)$ and $mH_1(i)$ are updated by the proposed algorithm by utilizing (6), the updated mass values of the SUs are sent to fusion center for final decision.

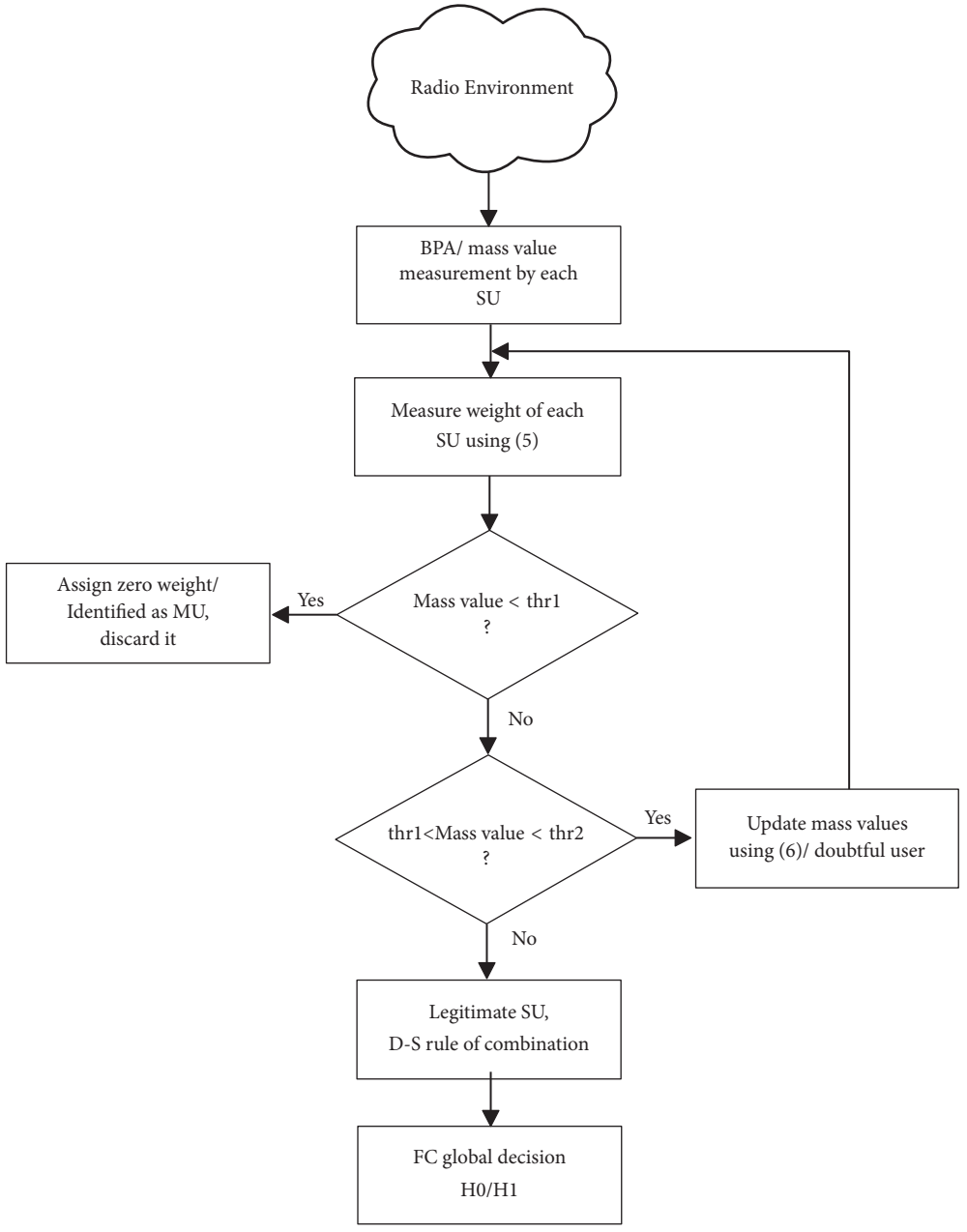


Figure 3: Flow chart of the proposed scheme.

According to the D-S evidence theory, the combination of updated masses at the fusion center can be given as

$$\begin{aligned} \mathrm{mH}_0 &= \mathrm{m}'\mathrm{H}_0\,(1) \oplus \mathrm{m}'\mathrm{H}_0\,(2) \oplus \ldots \mathrm{m}'\mathrm{H}_0\,(\mathrm{i}) \\ &= \frac{\sum_{A_1 \cap A_1 \cap \ldots A_N = H_0} \prod_{i=1}^{N} \mathrm{mH_i}\,(\mathrm{A_i})}{1-K} \\ \mathrm{mH}_1 &= \mathrm{m}'\mathrm{H}_1\,(1) \oplus \mathrm{m}'\mathrm{H}_1\,(2) \oplus \ldots \mathrm{m}'\mathrm{H}_1\,(\mathrm{i}) \\ &= \frac{\sum_{A_1 \cap A_1 \cap \ldots A_N = H_1} \prod_{i=1}^{N} \mathrm{mH_i}\,(\mathrm{A_i})}{1-K}, \end{aligned} \tag{9}$$

where $K = \sum_{A_1 \cap A_1 \cap \ldots A_N = \varnothing} \prod_{i=1}^{N} \mathrm{mH_i}(\mathrm{A_i})$ and the operator $\oplus$ is the sequential combination of the mass values.

The final decision F_d is determined based on the following simple strategy:

$$F_d = \begin{cases} \mathrm{H}_1\text{: } \mathrm{mH}_1 \succ \mathrm{mH}_0 \\ \mathrm{H}_0\text{: } \mathrm{mH}_1 \prec \mathrm{mH}_0. \end{cases} \tag{10}$$

The overall flowchart of the proposed scheme is given in Figure 3.

Input: energy of the sample (mass value $mH_0(i), mH_1(i)$), lowest threshold (thr_1), highest threshold (thr_2)
If energy of the sample (mass value) < lowest threshold (thr_1)
    Set the corresponding weights to zero
Else if lowest threshold (thr_1) < energy of the sample (mass value) < highest threshold (thr_2)
    Update masses value using (6)
Else
    Don't update mass values
End if
Output: Update mass value $m'H_0(i), m'H_1(i)$.

Algorithm 1: Proposed algorithm at fusion center.

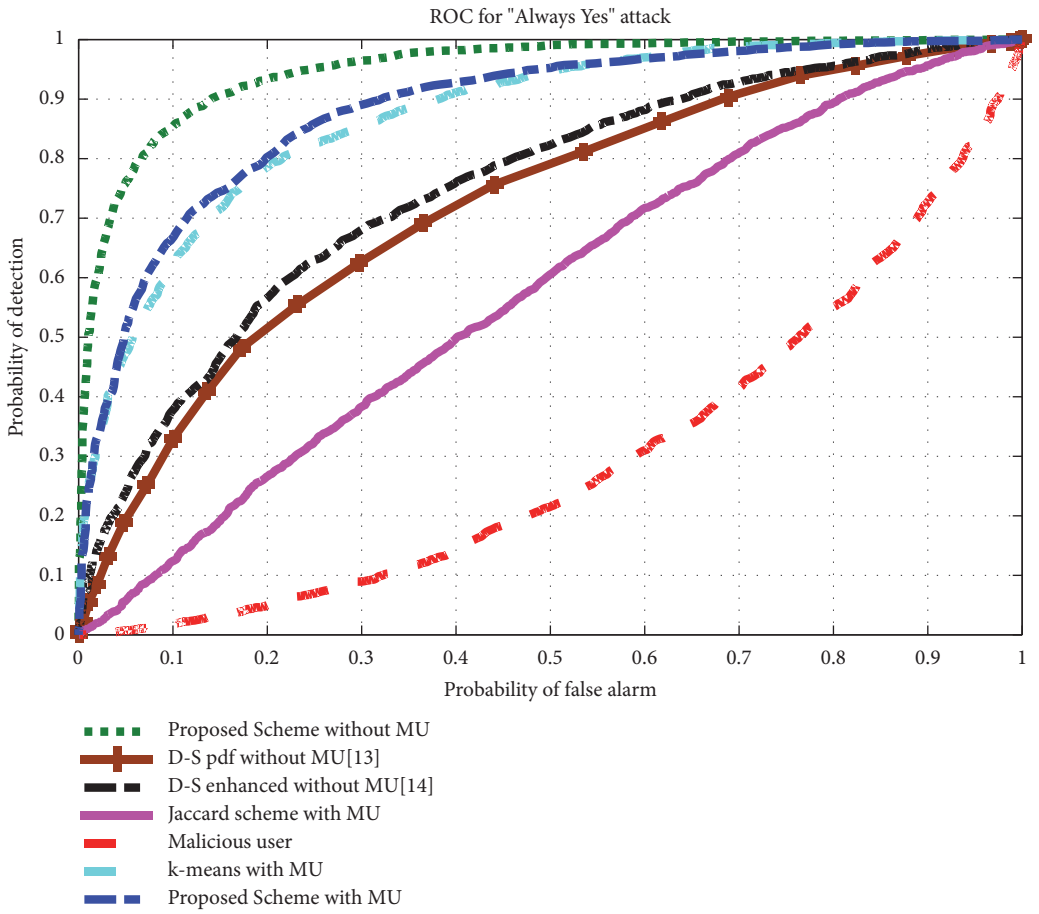


Figure 4: Performance comparison of various schemes with and without "Always Yes" attack.

4. Numerical Evaluation

In this section, we discuss the simulation results of the proposed scheme and compare its performance with the existing schemes. In the simulation environment, we placed five SUs randomly and measured their local energy by utilizing energy detector technique. The probability of appearance of the LU is 0.5, and the bandwidth is 6 MHz, and sensing period is 50 μsec. The simulation environment is developed by utilizing MATLAB as an implementation tool. The parameters for the simulations are summarized in Table 1.

In Figure 4, the performance comparison of the proposed scheme to other existing schemes is shown, with and without existence of MU in the network. In this scenario, we consider 20% malicious users for "Always Yes". It can be clearly observed that without malicious user in the network, for P_f of about 0.15, P_d of the proposed scheme is 0.9 when the highest P_d from all the other schemes is almost 0.85. Similarly, with the presence of MU in the network, P_d of the proposed scheme drops but it is still higher than other existing schemes. At P_f of 0.1, P_d of the proposed scheme

Table 1: Parameters of the simulation.

Parameters	Value
Number of SUs in the networks	5
Probability of appearance of LU	0.5
Proportion of MUs in the network	20%
Number of iteration	10,000

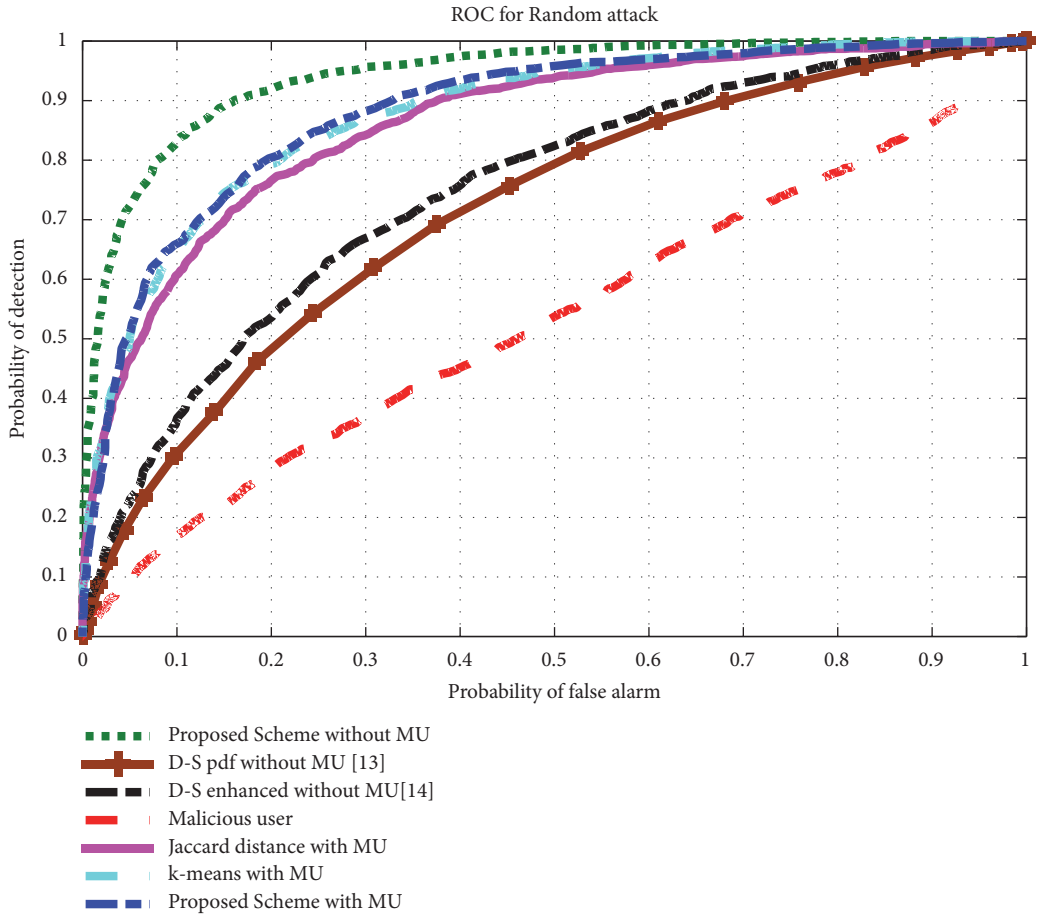


Figure 5: Performance comparison of various schemes with and without "Random" attack.

is almost 0.68 whereas the highest P_d of the other existing scheme is 0.64.

Figure 5 shows the performance comparison of the proposed scheme with other existing schemes with and without having malicious user in the system. The malicious users attack considered is "Random" attack scenario. It can be observed from Figure 5 that without any malicious user in the network, for P_f of about 0.1, P_d of the proposed scheme is 0.86 when the highest P_d from all the other schemes is almost 0.79. P_d is increased in case of a "Random" malicious user due to random behavior of the malicious user. When the malicious user is added to the network, P_d of the proposed scheme is still higher than the other existing schemes. At P_f of 0.1, P_d offered by the proposed scheme is almost 0.67 which is still higher than P_d of the other existing schemes.

Figure 6 shows the comparison of the proposed scheme with other existing schemes, with or without an "Always No" malicious user attack in the system. It is clear from Figure 6 that without any malicious user in the network, for P_f of about 0.2, P_d of the proposed scheme is almost 0.92 when the highest P_d from all the other schemes is almost 0.9. When the malicious user is added to the network, the performance of the proposed is better than the other existing schemes. Specifically, when the value of P_f is 0.2, P_d of the proposed scheme is almost 0.8 and still better than the other existing scheme.

5. Conclusion

The spectrum sensing data falsification attacks falsifies the sensing results, which highly degrades the performance of cooperative spectrum sensing. In this paper, we proposed a double adaptive approach in cognitive radio networks to deal with legitimate, doubtful, and malicious users in the

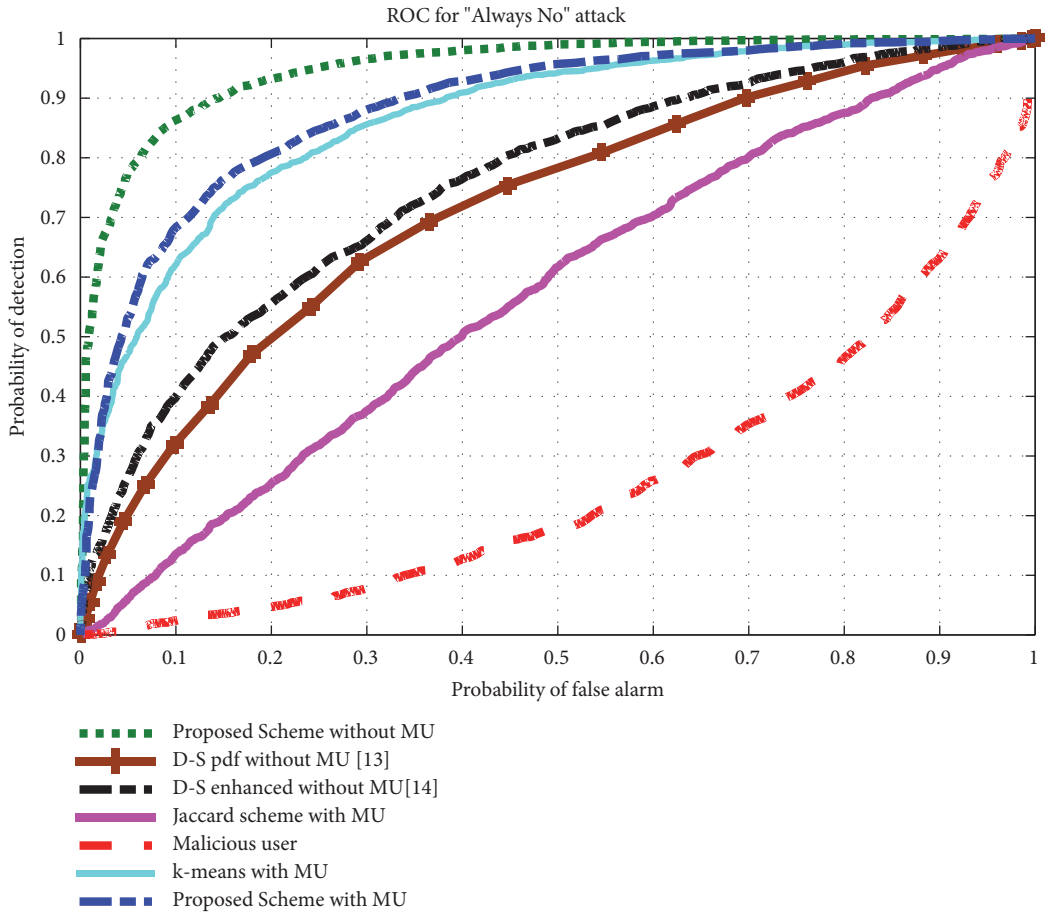


Figure 6: Performance comparison of various schemes with and without "Always No" attack.

networks. Maximal ratio combining scheme is utilized for weighting of the secondary users and the proposed double adaptive thresholding approach categorized legitimate, doubtful, and malicious users. A fair opportunity is provided to doubtful user to ensure its credibility. At the fusion center Dempster-Shafer evidence theory is utilized for combining legitimate secondary users' and making the final decision. The performance of the proposed scheme is tested in the presence of various types of malicious users' attacks and compared the results with the existing schemes. The results showed that the proposed scheme outperforms the existing schemes in case of "Always yes", "Always No", and Random attacks.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-0-01426) supervised by the IITP (Institute for Information and Communication Technology Promotion) and in part by the National Research Foundation (NRF) funded by the Korea government (MISP) (no. NRF-2017R1A2B1004474).

References

[1] M. Sansoy, A. S. Buttar, and K. Singh, "Cognitive Radio: Issues and Challenges," *Journal of Network Communications and Emerging Technologies (JNCET*, vol. 2, p. 4, 2015.

[2] FCC, *eT Docket No 03-222 Notice of Proposed Rule-Making And Order*, 2003.

[3] J. Mitola, "Cognitive Radio for Flexible Mobile Multimedia Communications," in *Proceedings of the IEEE International Workshop on Mobile Multimedia Communications*, 1999.

[4] *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio [Ph.D. thesis]*, KTH Royal Institute of Technology, 2000.

[5] A. Muralidharan, P. Venkateswaran, S. G. Ajay, D. A. Prakash, M. Arora, and S. Kirthiga, "An adaptive threshold method for energy based spectrum sensing in cognitive radio networks," in *Proceedings of the IEEE International Conference on control instrumentation communication and computational technologies (ICCICCT)*, 2015.

[6] P. Geete and M. Motta, "Analysis of Different Spectrum Sensing Techniques in Cognitive Radio Network," *International Research Journal of Engineering and Technology (IRJET)*, vol. 2, p. 5, 2015.

[7] M. Jenani, "Network Security, A Challenge," *International Journal of Advanced Networking and Application*, vol. 8, no. 5, pp. 120–123, 2017.

[8] J. Marinho, J. Granjal, and E. Monteiro, "A survey on security attacks and countermeasures with primary user detection in cognitive radio networks," *EURASIP Journal on Information Security*, vol. 4, p. 4, 2015.

[9] H. Du, S. Fu, and H. Chu, "A credibility-based defense ssdf attacks scheme for the expulsion of malicious users in cognitive radio," *International Journal of Hybrid Information Technology*, vol. 8, p. 9, 2015.

[10] A. Rauniyar and S. Y. Shin, "Cooperative adaptive threshold based energy and matched filter detector in cognitive radio networks," *Journal of Communication and Computer*, vol. 12, pp. 13–19, 2015.

[11] S. Mamatha and K. Aparna, "Design of an Adaptive Energy Detector based on Bi-Level Thresh holding in Cognitive Radio," *International Journal of Scientific Engineering and Technology Research*, vol. 4, no. 2, pp. 346–349, 2015.

[12] T. Peng, Y. Chen, J. Xiao, Y. Zheng, and J. Yang, "Improved soft fusion-based cooperative spectrum sensing defense against SSDF attacks," in *Proceedings of the 2016 International Conference on Computer, Information and Telecommunication Systems, CITS 2016*.

[13] Q. Peng, K. Zeng, W. Jun, and S. Li, "A distributed spectrum sensing scheme based on credibility and evidence theory," in *Proceeding of the IEEE 17th International symposium on Personal, Indoor and Mobile Radio Communication*, pp. 1–5, Helsinki, Finland, 2006.

[14] N. Nguyen-Thanh and I. Koo, "An enhanced cooperative spectrum sensing scheme based on evidence theory and reliability source evaluation in cognitive radio context," *IEEE Communications Letters*, vol. 13, no. 7, pp. 492–494, 2009.

[15] F. Ye, X. Zhang, and Y. Li, "Collaborative spectrum sensing algorithm based on exponential entropy in cognitive radio networks," *Symmetry*, vol. 9, p. 36, 2017.

[16] A. A. Sharifi and M. J. Musevi Niya, "Defense Against SSDF Attack in Cognitive Radio Networks: Attack-Aware Collaborative Spectrum Sensing Approach," *IEEE Communications Letters*, vol. 20, no. 1, pp. 93–96, 2016.

[17] N. Nguyen-Thanh and I. Koo, "A Robust Secure Cooperative Spectrum Sensing Scheme Based on Evidence Theory and Robust Statistics in Cognitive Radio," *IEICE Transactions on Communications*, vol. 92-B, no. 12, pp. 3644–3652, 2009.

[18] M. S. Khan and I. Koo, "The Effect of Multiple Energy Detector on Evidence Theory Based Cooperative Spectrum Sensing for Cognitive Radio Networks," *Journal of Information Processing Systems*, vol. 12, no. 2, pp. 295–309, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 2036818, 9 pages
https://doi.org/10.1155/2019/2036818

Research Article

Smart Real-Time Video Surveillance Platform for Drowsiness Detection Based on Eyelid Closure

Muhammad Tayab Khan,[1] Hafeez Anwar,[2] Farman Ullah,[2] Ata Ur Rehman,[2] Rehmat Ullah,[3] Asif Iqbal,[4] Bok-Hee Lee,[5] and Kyung Sup Kwak[4]

[1]*The Incubator, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi KPK, Pakistan*
[2]*Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Pakistan*
[3]*Department of Computer Systems Engineering, University of Engineering & Technology, Peshawar, Pakistan*
[4]*Department of Information and Communication Engineering, Inha University, Republic of Korea*
[5]*Department of Electrical Engineering, Inha University, Republic of Korea*

Correspondence should be addressed to Kyung Sup Kwak; kskwak@inha.ac.kr

Received 6 September 2018; Revised 23 November 2018; Accepted 22 January 2019; Published 19 March 2019

Guest Editor: Sarmadullah Khan

We propose drowsiness detection in real-time surveillance videos by determining if a person's eyes are open or closed. As a first step, the face of the subject is detected in the image. In the detected face, the eyes are localized and filtered with an extended Sobel operator to detect the curvature of the eyelids. Once the curves are detected, concavity is used to tell whether the eyelids are closed or open. Consequently, a concave upward curve means the eyelid is closed whereas a concave downwards curve means the eye is open. The proposed method is also implemented on hardware in order to be used in real-time scenarios, such as driver drowsiness detection. The evaluation of the proposed method used three image datasets, where images in the first dataset have a uniform background. The proposed method achieved classification accuracy of up to 95% on this dataset. Another benchmark dataset used has significant variations based on face deformations. With this dataset, our method achieved classification accuracy of 70%. A real-time video dataset of people driving the car was also used, where the proposed method achieved 95% accuracy, thus showing its feasibility for use in real-time scenarios.

1. Introduction

We propose an image-based framework for the detection and recognition of drowsiness based on a person's eyes. Such a framework can be instrumental in a multitude of scenarios, such as driver drowsiness detection, and thus has the potential to save lives. Among other things, the most common causes of driver drowsiness are fatigue and excessive alcohol consumption. In such cases, it is extremely important to detect the condition of the driver and take appropriate steps to save lives on the roads. Our proposed framework is a step towards a solution to this public issue. To that end, our framework continuously monitors a driver's condition in real time by using a video camera installed in front of the person. From the video, we use an image-based noninvasive technique to detect the eyes of the driver and classify them as being open or not.

However, detection and classification of a driver's eyes constitute a nontrivial problem that has a set of challenges. First is localization of the driver's face from among other passengers. We utilize the Viola-Jones [1] algorithm for face detection due to its real-time performance and robustness to scale and location variations. In addition to the driver's face, the algorithm detects all faces in a video that may contain some falsely detected faces. After localization of the candidate face, the next challenge is detection of the eyes. As with face detection, the Viola-Jones algorithm detects many regions in the face as eyes, among which are falsely detected eyes. Lastly, the major obstacle in our approach is the removal of eyebrow in the detected candidate eye region. We propose using the curvature of the eyelids to determine the state of the eye as being open or closed. However, the curvature of the eyebrows is also detected with our specially designed curvature detection filter.

We propose an incremental approach to solve these problems. As a first step, we detect faces with the Viola-Jones algorithm. The candidate face from among the detected faces is then decided as being the one that has the largest area assuming that the driver is nearest to the camera. Only this face region is then processed in subsequent video frames, reducing the processing cost. The eyes in this face region are then detected and processed for eyelid curvature detection. Finally, to detect the curvature of the eyelids, we apply a filter only to that part of the candidate eye region that is more likely to contain the eyelids. Consequently, noise induced by the eyebrows is reduced to a reasonable level. The curvature of the eyelid is a compact feature that is feasible for real-time scenarios due to its quick computation and acceptable accuracy.

1.1. Related Work. Most of the methods proposed for image-based drowsiness detection use symptoms related to the level of the driver's drowsiness. Detecting eyes that are closed or open is used in many methods. For instance, Dong and Wu [2] proposed detecting the eye state via distance calculation between the upper and lower eyelids. Dasgupta et al. [3] used the amount of eyelid closure as a cue for drowsiness. The presence of the iris in the image indicates that the eye is open. Detection of the iris in an eye image using the circular Hough transform is used in [4] to classify the eye as being open or closed. The ratio between the eye's height and width, as well as its area, is used in [5] to determine the status of the eye. Other methods include template matching [6, 7], use of local image features [8, 9], and using a Hidden Markov Model (HMM) [10]. Head pose estimation is used by Teyeb et al. [11], where the level of driver alertness is measured by the head being inclined to a certain degree and for a specific time period. A threshold placed on the changing rate of the mouth contour to detect yawning is used by Alioua et al. [12] to detect drowsiness of the driver. However, expert knowledge can be instrumental for accurate and timely detection of drowsiness. Such knowledge is implemented by Rezaei and Klette [13] using a fuzzy control fusion system. Intense devoted driving condition of such a driver who has already faced an accident before is called hypovigilance. This condition leads to rapid exhaustion and could cause drowsiness. Smith et al. [14] proposed using a finite state machine (FSM) to detect hypovigilance and used it as a cue for drowsiness detection.

In addition to these techniques, specialized machine learning-based methods have been proposed with the recent emergence of deep learning. Park et al. [15] proposed a deep architecture, referred to as deep drowsiness detection (DDD). It is reliable for the exclusion of backgrounds and environment variations and achieved 73.06% accuracy. Weng et al. [16] introduced a novel hierarchical temporal deep belief network (HTDBN) for drowsiness detection. Their work highlights the detection of head positions and faces to detect drowsiness. Huynh et al. [17] used a three-dimensional convolutional neural network (CNN) to extract features in spatial-temporal domain. The method is designed to solve issues with extreme head poses, where it achieved 87.46% accuracy. Shih and Hsu [18] use a multistage spatial-temporal network (MSTN) with a CNN to detect various states of drowsiness. They achieved 82.61% accuracy. Lyu et al. [19] used random forest method to extract effective facial descriptors to describe drowsiness based on face alignment, and to classify the driver's facial states where the claimed accuracy is 88.18%. Although, a deep learning-based method achieves state-of-the-art performance, success still lies in the availability of large amounts of data to train deep nets. In contrast, our proposed method is purely based on image processing techniques that are suitable for real-time implementation. Our method is a stepwise procedure to determine the drowsiness of a driver. Following are the major steps:

(1) face detection in the image,

(2) driver's face extraction among the detected faces,

(3) extraction of the region of interest (ROI) that contains the driver's face,

(4) eye extraction and eyebrow removal from the detected driver's face,

(5) extraction of the eyelids from the detected eyes,

(6) determination of the eyelids as being concave up or down and classification of the eyes based on this concavity as being open or closed.

Explanation for each of the steps follows.

1.2. Faces Detection in the Image. Faces in the scene are detected via the famous Viola-Jones object detection algorithm [1] due to its fast calculation time. Following are the main steps of this algorithm:

(1) integral image calculation,

(2) feature detection,

(3) AdaBoosting for redundant features rejection,

(4) classification of the detected features using cascade of classifiers.

1.3. Driver's Face Extraction among the Detected Faces. The Viola-Jones framework is likely to detect multiple faces in the image as shown in Figure 1. Due to this reason, the major issue now is to extract the face of interest which is the driver's face. In our proposed setup, the camera is installed closer to the driving seat. Consequently, the face of the driver will occupy more number of pixels in the image than any other detected faces. From Figure 1 it can be observed that among all those detected faces, the driver's face has the highest area in the image.

1.4. Extraction of the Region of Interest (ROI) That Contains the Driver's Face. For fast processing as per demand of real-time application, we do not need to process the whole image in the consecutive frames but to process only the region of interest (ROI) that contains the driver's face. The parameters of the ROI are taken from the very first frame and then used in the rest of the driving session. The ROI is taken as the extension

Figure 1: Multiple faces detection in a driving-like scenario.

of all four sides of driver's face with 80% of its width as shown in Figure 2. In subsequent frames, we extract the ROI for the processing of driver's face, thus reducing the computational cost and time.

1.5. Eye Extraction and Eyebrow Removal from the Detected Driver's Face. The next step after face detection is to extract the driver's eyes from the detected face image. However, the bounding boxes for the eyes also contain eyebrows which act as noise, as shown in Figure 3. Since, our proposed method is based on the curvature of the eyelids and the eyebrows have a similar orientation, they are likely to generate falsely detected eyelids. Therefore, the eyebrows must be removed from the detected eye bounding boxes in order to achieve accurate eyelid detection. We propose combining the regions in the bounding boxes of both eyes in order to remove the eyebrows. To this end, the extreme points of both the bounding boxes are considered. Let $S_L(x_1, y_1)$ and $S_R(x_1, y_1)$ be the top left extreme points of both bounding boxes, as shown in Figure 3(a). Similarly, $E_L(x_2, y_2)$ and $E_R(x_2, y_2)$ are the bottom right extreme points of the left and right bounding boxes, respectively. Since, the combined bounding box for both eyes is the combination of both these bounding boxes, the extreme points of the combined bounding box are determined with the help of the extreme points of the individual bounding boxes. First, the x-coordinate of the upper left corner of the combined bounding box labeled as S_{Box}, as shown in Figure 3(b), is determined. Since $S_L(x_1) < S_R(x_1)$, the x-coordinate of S_{Box} is equal to $S_L(x_1)$. Similarly, S_R is lower than S_L along the y-axis, i.e., $S_R(y_1) > S_L(y_1)$, so we take the y-coordinate of S_{Box} equal to $S_R(y_1)$. The bottom right extreme point, i.e., E_{Box} of the combined region, is determined in a similar manner. Combining the bounding boxes of both the left and right bounding boxes helps to eliminate regions that contain eyebrows. Nonetheless, this combined region still contains a considerable part of the eyebrows, as shown in Figure 3(b). We propose eliminating the eyebrows by lowering the top left corner of the combined region by 20% of the height of the region along the y-axis as shown in Figure 3(c).

1.6. Extraction of the Eyelids from the Detected Eyes. The eyelids of the detected eyes are extracted via their curvature. We assume that the curvature of the eyelid's edge is greater for open eyes than for closed eyes. However, before the measurement of the curvature, the edge of the eyelid should be detected. We propose modifying the Sobel operator so it can detect the curved edges in the detected eyes. This is shown in Figure 4, along with its response on an open eye. To get a pronounced response and to reject extra curvature detection like that of the iris, the filter is applied again to its own response. This is shown in Figure 5, where the first and second responses of the filter are depicted. Several contours are detected by the double response of the filter, where we consider the contour with the largest area as the eyelid.

1.7. Eye Classification Based on Eyelid Curvature. Once the eyelid is detected, the next step is to classify the curve as being upward or downward curve in order to determine the state of the eye. Concavity of the detected eyelid curve is determined with the help of its two extreme points and their mid points along both the x-axis and y-axis. If the curve of the eyelid is "concave down" then the eye is "open"; however, if the curve is "concave up" then the eye is "closed". The concavity approximation of the eyelid curve is explained as follows.

Let the two extreme points of the curve be $A(x_1, y_1)$ and $B(x_2, y_2)$, as shown in Figure 6 for both concave up and concave down curves. Let the line segment passing through the mid of these extreme points with respect to their y coordinates be $E'E''$, as shown in Figures 6(a) and 6(d). Then, the y intercept of this line is given as follows.

$$E = \left| \frac{y_2 - y_1}{2} \right| \tag{1}$$

Similarly, the line segment that passes through the mid of the extreme points with respect to their x coordinates is $C'C''$, as shown in Figures 6(b) and 6(e). The x intercept of the line $C'C''$ is given as follows.

$$C = \frac{x_2 - x_1}{2} \tag{2}$$

This line intersects the eyelid curve at point D. Let line segment $D'D''$ that is parallel to the x-axis intersects the eyelid curve at point D as shown in Figures 6(b) and 6(e). Consequently, we get two "*curve handling*" line segments, namely, $E'E''$ and $D'D''$, that can determine the concavity of the eyelid curve. We can see in Figures 6(c) and 6(f) that,

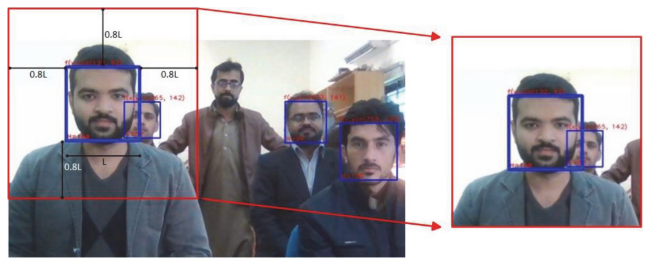
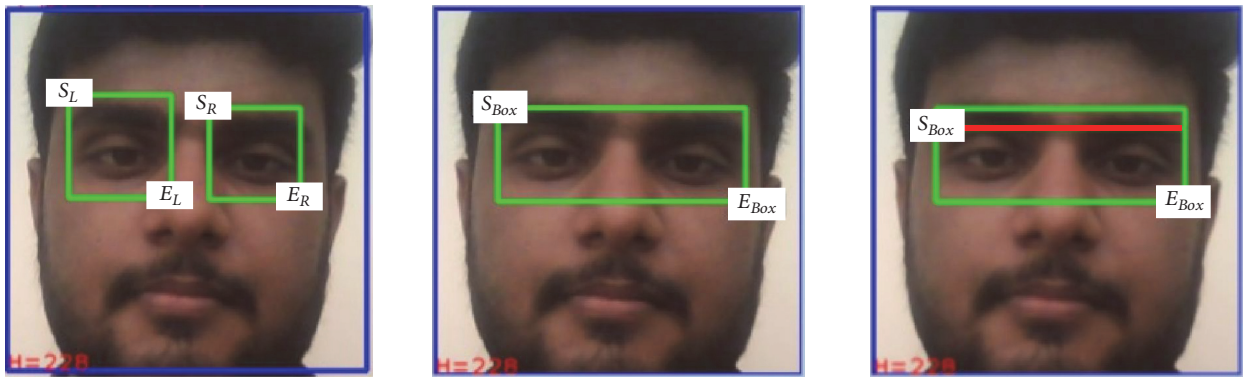


Figure 2: Region of interest (ROI) depicting driver's face.

(a) (b) (c)



Top left corner of left bounding box: $S_L(x_1,y_1)$

Top left corner of right bounding box: $S_R(x_1,y_1)$

Top left corner of combined bounding box: S_{Box}

Bottom right corner of left bounding box: $E_L(x_2,y_2)$

Bottom right corner of right bounding box: $E_R(x_2,y_2)$

Bottom right corner of combined bounding box: E_{Box}

Figure 3: Removal of eyebrows.

$$F = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & -2 & 0 & -1 & 0 \end{bmatrix}$$

(a) (b) (c)

Figure 4: (a) Original image of the eye, (b) the proposed filter, and (c) response of the filter on eye image.

(a) (b) (c)

Figure 5: (a) Original image of the eye, (b) first response of the filter, and (c) second response of the filter applied to the first response.

if line segment $E'E''$ lies above $D'D''$ along the y-axis, then the curve is concave up, and the eye is closed. If line segment $D'D''$ lies above $E'E''$ along y-axis, then the curve is concave down and the eye is open.

However, such ideal curves are not always detected in a real scenario. Instead, half curves are mostly detected due to shade of the nose over the eyes which results in a low contrast region in the image. Consequently, our proposed curve approximation method also works for such a condition which is shown in Figure 7.

2. Real-Time Drowsiness Detection and an Alarm Generation System

The proposed method is used to design an alarm generation system for the alertness of drowsy drivers. This system is built

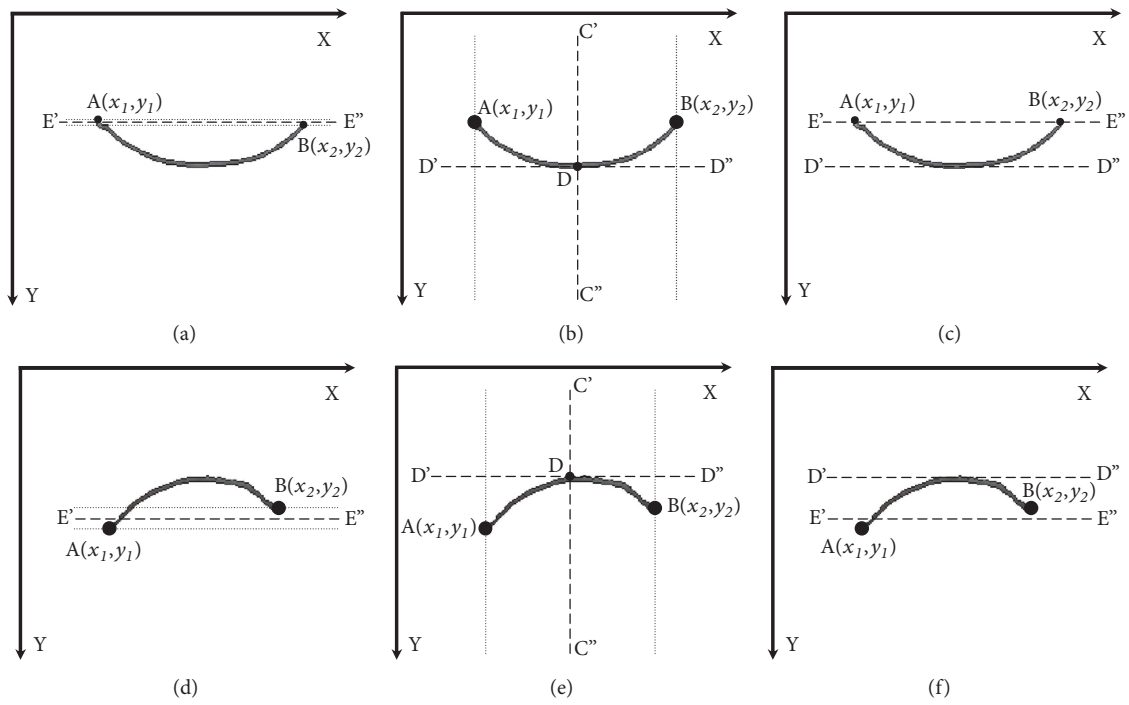


Figure 6: (a)-(c) Ideal eyelid curve for open eye, (d)-(f) ideal eyelid curve for closed eye.

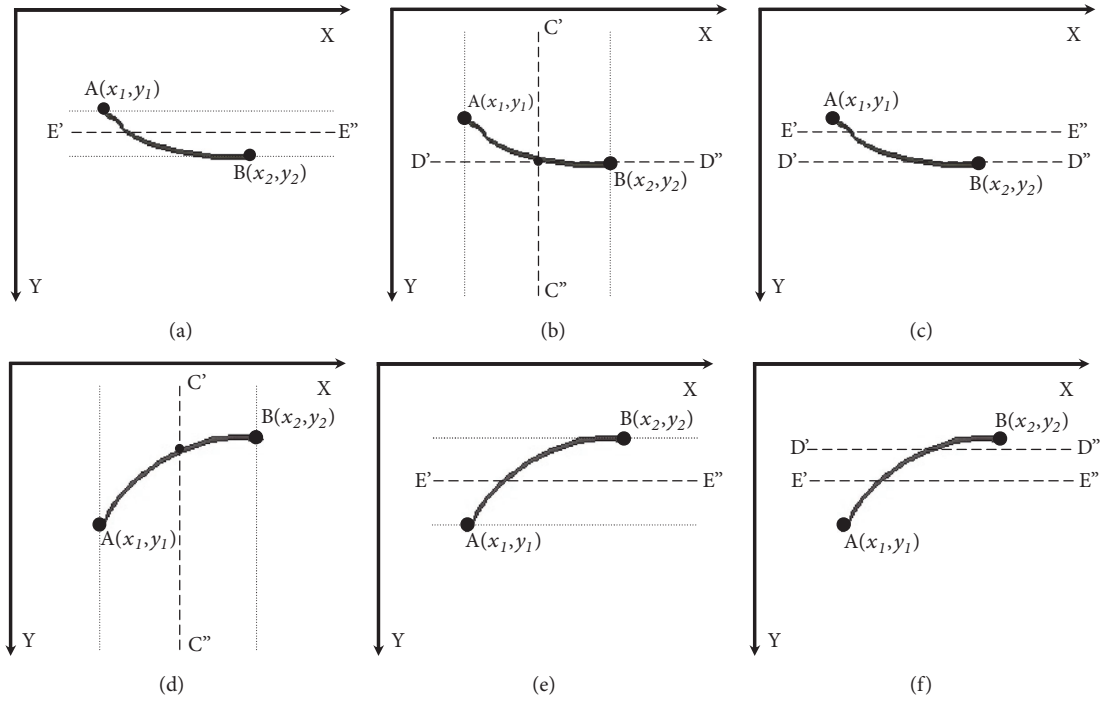


Figure 7: Ideal curves of the phenomenon when the eyelid curve is half detected.

on a Raspberry Pi with an interfaced camera for real-time video capture. A buzzer and a light-emitting diode (LED) are also interfaced to the Raspberry Pi to generate an alarm if the driver is detected as drowsy. The whole system is shown in Figure 8. A flow chart of the algorithm implemented in the system is shown in Figure 9. It consists of two main blocks. The first is the preprocessing block where the face detection is performed on the real-time video captured by the camera. Once a face is found, its ROI is detected and processed in the second block to detect driver drowsiness via eyelid closure. The eyes of the driver are continuously monitored, and if found closed for a certain period of time, the alarm sounds.

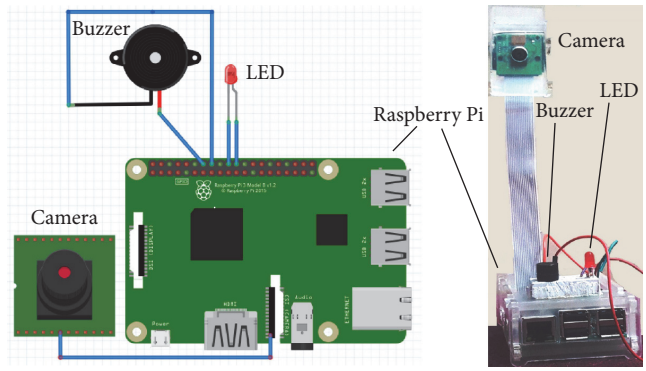


Figure 8: Real-time system for driver drowsiness detection and alarm generation.

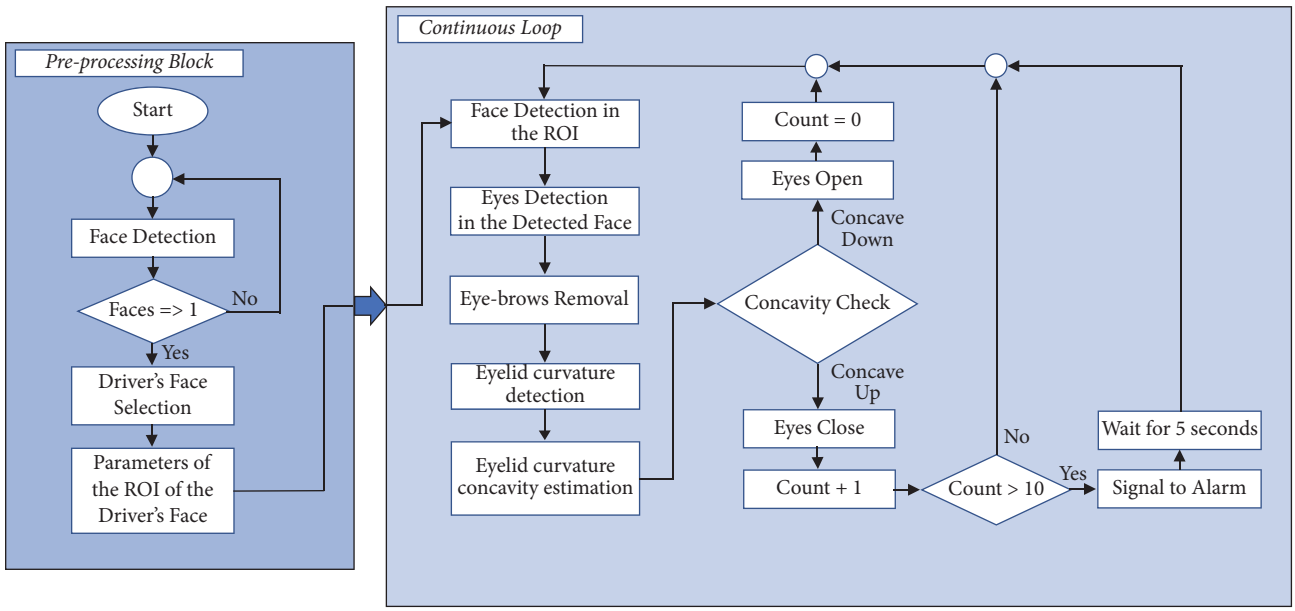


Figure 9: Flow chart of the real-time drowsiness detection and alarm generation system.

3. Results and Discussion

In order to evaluate our proposed method, we used image datasets that are incrementally difficult. For instance, as a first step, we generated our own image dataset where a person's face is imaged with uniform background. We acquired images of 319 persons with their eyes open and closed. Detailed statistics are shown in Table 1, where we outlined the accuracy of face detection, eye detection, and then classification of the eyes as open or closed. The detection results of both the open and closed eyes are shown in Figure 10. The Viola-Jones algorithm gives 100% face detection accuracy on our not-so challenging dataset due to the absence of background clutter and the uniform illumination from an in-door environment. Consequently, the eyes are detected with almost perfect accuracy (i.e., 98%). Our proposed algorithm for eyelid curvature detection and eye classification based on this detection gives almost 95% accuracy. However, this accuracy is dependent upon the eyes detection accuracy which is 98%. Therefore, the relative accuracy of our algorithm becomes 97%.

We also evaluated our proposed method on a bench mark dataset [20]. The images in the dataset are challenging due to image variations caused by face deformations, out-of-plane orientations, glasses, and irregular illuminations. Since, our basic assumption is that the subject will be looking towards the camera, we rejected images in the dataset with severe out-of-plane and in-plane rotations. The size of the images is 100×100 that are scaled up to 150×150 in order to meet the requirements of our proposed method. Table 2 shows the statistics achieved by the proposed method on the dataset. Some of the correct eye classification results are shown in Figure 11 while incorrect results are shown in Figure 12.

Lastly, the proposed method was evaluated with a real-time video of someone driving a car. The video was recorded in the day time with variations in illumination due to differences in the direction of sunlight. Two videos were recorded to evaluate the proposed method, for which statistics are shown in Table 3. Some exemplar frames where the eyes are correctly and wrongly classified are shown in Figures 13 and 14, respectively. The face and eye detection accuracies

Figure 10: Detection results of open and closed eyes from images with a homogeneous background.

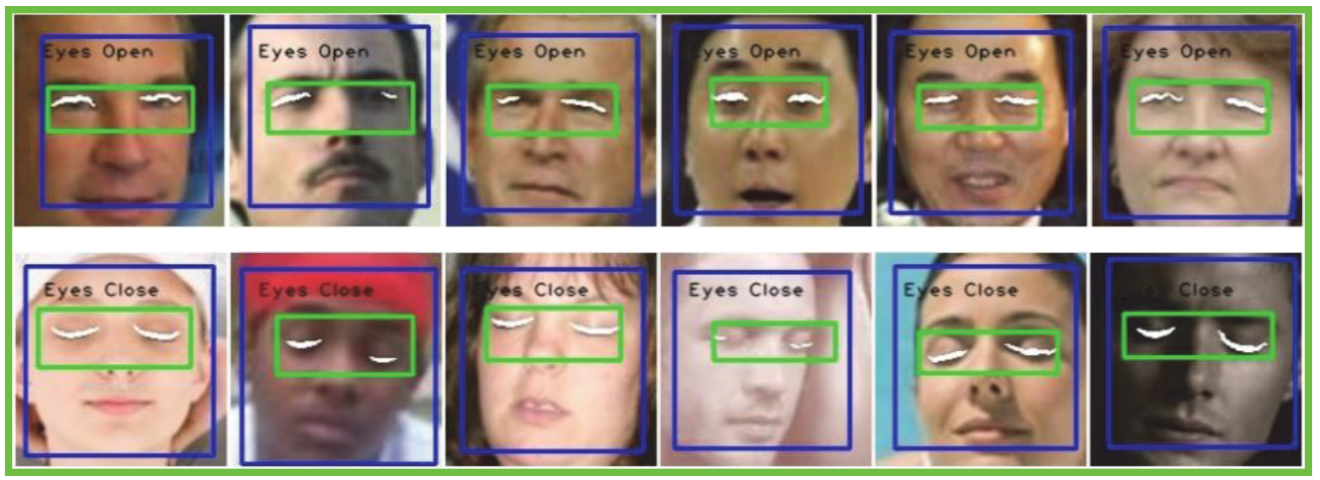


Figure 11: Correct detection of open and closed eyes on the bench mark dataset [20].

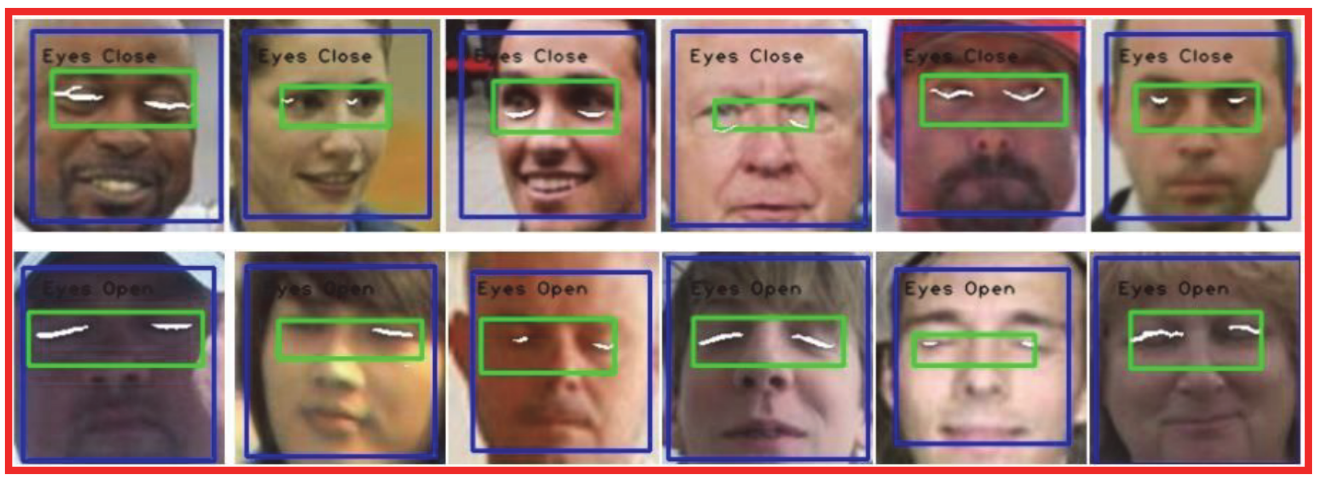


Figure 12: Incorrect detection of open and closed eyes from the bench mark dataset [20].

Table 1: Open and closed eye detection from images with uniform background.

Type of dataset	Number of images	Face Detection Accuracy	Eye Detection Accuracy	Eye Classification Accuracy
Close Eyes Dataset	319	100%	97.43%	90.31%
Open Eyes Dataset	319	100%	98.73%	99.26%
Total	638	100%	98.08%	94.78%

Table 2: Open and closed eye detection rates from images in a benchmark dataset [20].

Type of dataset	Number of images	Face Detection Accuracy	Eye Detection Accuracy	Eye Classification Accuracy
Close Eyes Dataset	869	90.45%	73.80%	68.70%
Open Eyes Dataset	807	98.80%	98.37%	73.17%

Figure 13: Correct detection of open and closed eyes on the video dataset.

Figure 14: Incorrect detection of open and closed eyes on the video dataset.

from both the videos are not encouraging. However, the main purpose of the proposed algorithm is to classify eyes in the detected face and eye images. To this end, it achieves over 95% eye classification accuracy from both videos. This shows the feasibility of the proposed method for driver drowsiness detection. It should be noted, however, that these results were achieved during the day time. For night time, our proposed method can be used on top of face and eye detection algorithms that work during night time.

4. Conclusion

A method for image-based drowsiness detection in real-time driving surveillance videos is proposed. It is a four-step method that first detects the face of the driver in the image from among several detected faces. Secondly, it extracts the eyes from the detected faces. In the third step, the curvature of the eyelids is detected using a modified Sobel operator. Finally, the eyes are classified as closed or open based on the curvature of the eyelids. The proposed method achieved an average classification accuracy of 95% on a simple image dataset with homogeneous backgrounds, an average classification accuracy of 70% on a complex benchmark image dataset, and greater than 95% classification accuracy on two real-time driving surveillance videos. However, the proposed method works only in the day time; its adaptation to night time will be explored in future work with more state-of-the-art face and eye detection algorithms. Similarly, more challenging face images where subjects might have glasses or phones will be used to evaluate the proposed method.

Table 3: Classification accuracy achieved on real-time video datasets.

Type of Dataset	Number of Frames	Face Detection Accuracy	Eye Detection Accuracy	Classification Accuracy
Closed Eyes video 1	1839	73.57%	73.17%	95.00%
Closed Eyes video 2	1387	96.75%	67.00%	95.00%
Open Eyes video 1	3601	44.32%	70.30%	90.50%
Open Eyes video 2	3201	78.01%	96.19%	95.20%
Total	10028	73.21%	76.66%	93.79%

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.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT)-NRF-2017R1A2B2012337.

References

[1] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[2] W. Dong and X. Wu, "Driver fatigue detection based on the distance of eyelid," in *Proceedings of the 2005 IEEE International Workshop on VLSI Design and Video Technology, IWVDVT 2005*, IEEE, pp. 365–368, China, May 2005.

[3] A. Dasgupta, A. George, S. L. Happy, and A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1825–1838, 2013.

[4] N. Alioua, A. Amine, M. Rziza, and D. Aboutajdine, "Eye state analysis using iris detection based on Circular Hough Transform," in *Proceedings of the 2011 International Conference on Multimedia Computing and Systems, ICMCS'11*, IEEE, pp. 1–5, Morocco, April 2011.

[5] Y. Du, P. Ma, X. Su, and Y. Zhang, "Driver fatigue detection based on eye state analysis," in *Proceedings of the 11th Joint Conference on Information Sciences*, 2008.

[6] X. Liu, X. Tan, and S. Chen, "Eyes closeness detection using appearance based methods," in *Proceedings of the International Conference on Intelligent Information Processing*, vol. 385, pp. 398–408, Springer, 2012.

[7] G. Fa-deng and H. Min-xian, "Study on the detection of locomotive driver fatigue based on image," in *Proceedings of the 2010 2nd International Conference on Computer Engineering and Technology*, pp. V7-612–V7-615, Chengdu, China, April 2010.

[8] L. Zhou and H. Wang, "Open/closed eye recognition by local binary increasing intensity patterns," in *Proceedings of the 2011 IEEE 5th International Conference on Robotics, Automation and Mechatronics, RAM 2011*, IEEE, pp. 7–11, China, September 2011.

[9] M. Tafreshi and A. M. Fotouhi, "A fast and accurate algorithm for eye opening or closing detection based on local maximum vertical derivative pattern," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 6, pp. 5124–5134, 2016.

[10] H. Qin, J. Liu, and T. Hong, "An eye state identification method based on the embedded hidden Markov model," in *Proceedings of the 2012 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2012*, IEEE, pp. 255–260, Turkey, July 2012.

[11] I. Teyeb, O. Jemai, M. Zaied, and C. B. Amar, "A drowsy driver detection system based on a new method of head posture estimation," in *Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning*, pp. 362–369, Springer, 2014.

[12] N. Alioua, A. Amine, and M. Rziza, "Driver's fatigue detection based on yawning extraction," *International Journal of Vehicular Technology*, vol. 2014, Article ID 678786, 7 pages, 2014.

[13] M. Rezaei and R. Klette, "Look at the driver, look at the road: No distraction! No accident!," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 129–136, USA, June 2014.

[14] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 4, pp. 205–218, 2003.

[15] S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks," in *Proceedings of the Computer Vision – ACCV 2016 Workshops*, vol. 10118 of *Lecture Notes in Computer Science*, pp. 154–164, Springer International Publishing, Cham, Switzerland, 2016.

[16] C. Weng, Y. Lai, and S. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Proceedings of the Computer Vision – ACCV 2016 Workshops*, vol. 10118 of *Lecture Notes in Computer Science*, pp. 117–133, Springer International Publishing, Cham, Switzerland, 2016.

[17] X.-P. Huynh, S.-M. Park, and Y.-G Kim, "Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine," in *Proceedings of the Computer Vision – ACCV 2016 Workshops*, vol. 10118 of *Lecture Notes in Computer Science*, pp. 134–145, Springer International Publishing, Cham, Switzerland, 2016.

[18] T. Shih and C. Hsu, "MSTN: multistage spatial-temporal network for driver drowsiness detection," in *Proceedings of the Computer Vision – ACCV 2016 Workshops*, vol. 10118 of *Lecture Notes in Computer Science*, pp. 146–153, Springer International Publishing, Cham, Switzerland, 2016.

[19] J. Lyu, H. Zhang, and Z. Yuan, "Joint shape and local appearance features for real-time driver drowsiness detection," in *Proceedings of the Computer Vision – ACCV 2016 Workshops*, vol. 10118 of *Lecture Notes in Computer Science*, pp. 178–194, Springer International Publishing, Cham, Switzerland, 2016.

[20] F. Y. Song, X. Y. Tan, X. Liu, and S. C. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," *Pattern Recognition*, vol. 47, no. 9, pp. 2825–2838, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 6470359, 20 pages
https://doi.org/10.1155/2019/6470359

Review Article

Underwater Wireless Sensor Networks: A Review of Recent Issues and Challenges

Khalid Mahmood Awan,[1] Peer Azmat Shah,[1] Khalid Iqbal,[2] Saira Gillani,[3] Waqas Ahmad,[1] and Yunyoung Nam[4]

[1]*Internet Communication & Networks (ICNet) Research Lab, Department of Computer Science, COMSATS University Islamabad, Attock 43600, Pakistan*
[2]*Pattern Recognition, Images and Data Engineering (PRIDE) Lab, Department of Computer Science, COMSATS University Islamabad, Attock 43600, Pakistan*
[3]*College of computing and informatics Saudi Electronic University, Jeddah, Saudi Arabia*
[4]*Department of Computer Science and Engineering, Soonchunhyang University, Asan 31538, Republic of Korea*

Correspondence should be addressed to Yunyoung Nam; ynam.sch@gmail.com

Received 29 June 2018; Revised 1 October 2018; Accepted 11 November 2018; Published 1 January 2019

Guest Editor: Sungchang Lee

Underwater Wireless Sensor Networks (UWSNs) contain several components such as vehicles and sensors that are deployed in a specific acoustic area to perform collaborative monitoring and data collection tasks. These networks are used interactively between different nodes and ground-based stations. Presently, UWSNs face issues and challenges regarding limited bandwidth, high propagation delay, 3D topology, media access control, routing, resource utilization, and power constraints. In the last few decades, research community provided different methodologies to overcome these issues and challenges; however, some of them are still open for research due to variable characteristics of underwater environment. In this paper, a survey of UWSN regarding underwater communication channel, environmental factors, localization, media access control, routing protocols, and effect of packet size on communication is conducted. We compared presently available methodologies and discussed their pros and cons to highlight new directions of research for further improvement in underwater sensor networks.

1. Introduction

Technique of sending and receiving message under the utilization of sound propagation in underwater environment is known as acoustic communication. Underwater sensor networks have number of vehicles and sensors that deploy in a specific area to perform collaborative monitoring and data collection tasks [1]. Traditionally for the monitoring of ocean bottom, oceanographic sensors are deployed for recording data at a fix location and recover the instruments at the completion of task. The major disadvantage of traditional approach is lack of interactive communication between different ends, recorded data can never get during any mission, and in case of any failure recorded data will be destroyed.

Underwater Sensor Networks support a wide variety of applications [2]; for example, aquatic surveillance, river and sea pollution discovery, monitoring, oceanographic data compilation, and commercial exploit the aquatic environment [3]. Underwater Sensor Networks can be utilized in any scenario from underwater warfare to the monitoring of environmental conditions [2]. Underwater Sensor Networks face constraints like limited bandwidth, high propagation delay, 3D topology, and power constraints. Radio and optical waves are not feasible for communication at each point of ocean. Under the entire limitations underwater sensor networks can only utilize acoustic signal that is a technique which is utilized by nature from the birth of ocean [4, 5]. Speed of sound is considered constant in underwater environment. However, speed of sound is affected by temperature, depth, and salinity of underwater environment. These factors produce variations in speed of sound in underwater environment [6]. Underwater acoustic channel frequencies spectrum, especially on mid-frequencies, is heavily shared

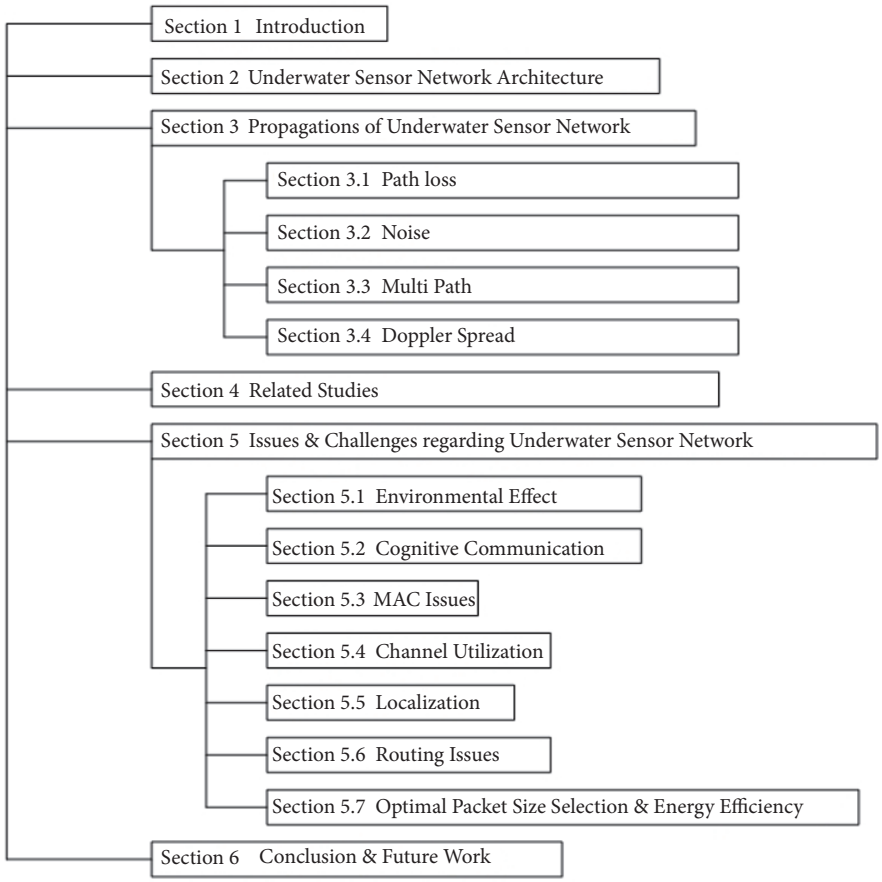


Figure 1: Overview of research work.

by various acoustic users in underwater environment. Still acoustic spectrum is temporally and spatially underutilized in underwater environment [7]. Variable characteristics of underwater environment have become a challenge for utilizing acoustic channel. For example, multipath propagation results in fading and phase fluctuations; Doppler Effect is observed due to the movement of both the sender and receiver nods. Speed of sound and underwater noise are other factors that influences the performance of acoustic channel [8].

Underwater sensor networks nodes are not static like ground-based sensor networks nodes. Instead, they move due to different activities and circumstances of underwater environment, usually 2-3m/sec with water currents. Sensed data is meaningful only when localization is involved. Another major issue that is affecting underwater sensor networks is energy saving. Because of nodes mobility, the majority of offered energy competent protocols become inappropriate for underwater sensor networks. Different protocols regarding land-based sensor networks are, for example, Directed Diffusion, Gradient, Rumor routing, TTDD, and SPIN. However, because of mobility and rapid change in network topology these existing grounds based routing protocols cannot perform efficiently in underwater environment [9]. Optimal packet size is depending on protocol characteristic like offered load and bit error rate. Poor packet size selection decreases the performance of the network throughput efficiency, latency, and resource utilization and energy consumption in multihop underwater networks can be greatly improved by a using optimum packet size [10–13].

To improve the better utilization of the available resources in underwater environment considering the energy and life time of network is discussed in detail in this paper. Balancing of energy consumption is carried out in underwater environment using the proposed techniques. The important contributions of this work are not only to highlight the deep and shallow ocean characteristics, but also to present the effect of temperature in acoustic communication and effect of temperature in noise, errors and protocols due to variation in environmental factors. In addition, classification of routing protocols for UWSNs and their comparison in terms of bounded latency, multipath, load balancing, energy consumption, geographic information, communication overhead, and time complexity. Similarly, data delivery ratios for single and multipath and the strengths and weaknesses of MAC protocols, with the used topology, are compared [14–16].

The paper is organized as illustrated in Figure 1. Section 2 presents the architecture of Underwater Wireless Sensor Networks. Section 3 describes propagation phenomena of

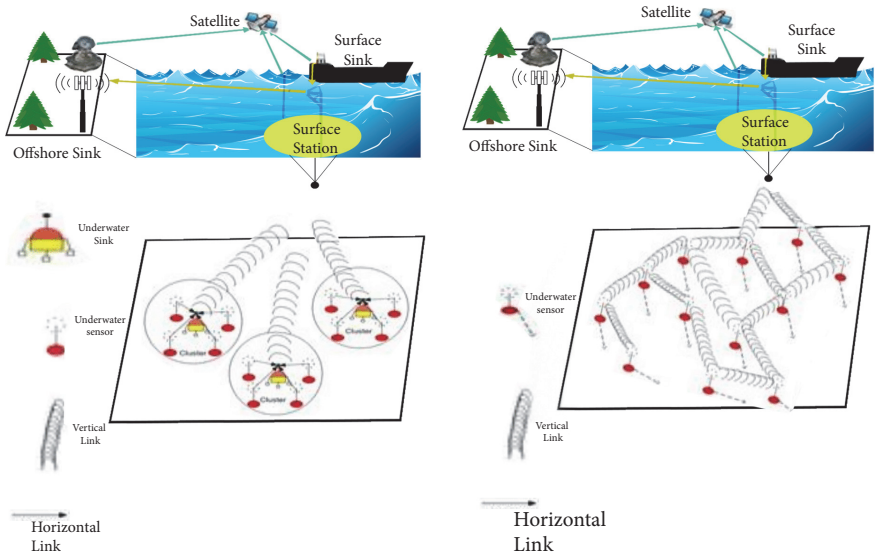


Figure 2: Two- and 3-dimensional networks architecture for UWSN regarding communication given in (a) and (b), respectively.

Underwater Wireless Sensor Networks. Section 4 presents previous achievements of different researcher in the form of related studies. The issues and challenges regarding underwater sensor networks are described in Section 5. Conclusion and future work are made in Section 6.

2. Underwater Sensor Networks Architecture

Underwater network's physical layer utilizes acoustic technology for communication. Limited bandwidth, capacity, and variable delays are characteristics of acoustic technology. Therefore, new data communication techniques and efficient protocols are required, for underwater acoustic networks. Designing the network topology requires significant devotion from designer, because underwater network performance is generally depending upon topology design. Network reliability should increase with efficient network topology and network reliability should also decrease with less efficient topology. Energy consumption of efficient network topology is highly less as compared to incorrect and less efficient topology design of underwater network. Design of topology for underwater sensor network is an open area for research [18, 19]. Underwater sensor networks architecture is shown in Figure 2.

2.1. Underwater Sensor Networks in Two-Dimensions. Deep ocean anchors are utilized for collection of sensor nodes in two-dimensional underwater sensor network architecture. Anchored underwater nodes use acoustic links to communicate with each other or underwater sinks. Underwater sinks are responsible to collect data from deep ocean sensors and provide it to offshore command stations, using surface stations. For this purpose, underwater sinks are provided in the company of horizontal and vertical acoustic transceivers. Purpose of horizontal transceivers is to communicate with sensor node, to collect data or provide them commands, as have been received by offshore command station, although vertical transceiver is used to send data to command station. Because ocean can be as deep as 10 km, vertical transceiver should contain enough range. Surface sink that is equipped with acoustic transceivers has the capability to manage parallel communication, by means of multiple organized underwater sinks. Surface sink is also equipped through extensive range radio frequency transmitters, to communicate with offshore sinks [18–21].

2.2. Underwater Sensor Networks in Three Dimensions. Activity required to present three-dimensional environments new architecture which is known as underwater three-dimensional networks is used. Sensor nodes float at different depth to monitor a specific activity in three-dimensional underwater networks. Traditional solution regarding underwater three-dimensional sensor networks is the use of surface buoys that provide ease in deploying such kind of network. But this solution is vulnerable to weather and tampering. Also, effortlessly can be discovered and disabled by enemies in the scenario of military operation. In underwater three-dimensional sensor networks architecture, ocean bottom is utilized to anchored sensor nodes. Depth of these nodes is controlled using wires which are attached with these anchors. Major challenge regarding such network is influenced by the current properties of the oceans [18–20].

3. Propagation Phenomena of Underwater Sensor Networks

Acoustic communication regarding underwater environment is a complex phenomenon because a lot of environmental factors affect acoustic communication. These factors are variable like long propagation delays, environmental noise,

Table 1: Deep and shallow ocean characteristics.

Characteristics	Shallow Ocean	Deep Ocean
Depth	0 m to 100 m	100 m to 10000 m
Temperature	High	Low
Multi-path Loss	Surface Reflection	Both Surface and Bottom Reflection
Spreading Factor (K)	Cylindrical	Spherical

path loss, Doppler spread, and multipath effect. Underwater environmental factors make acoustic channel highly variable. They also create bandwidth dependency upon both frequency and distance between two nodes. Generally, ocean is divided into two parts; these are shallow and deep ocean. Shallow and deep ocean characteristics are described in the Table 1. Shallow ocean highly affects acoustic channel because of high temperature gradient, multipath effect, surface noise, and large propagation delays, as compared to deep ocean. Underwater environment major propagation factors that affect acoustic communication are described in subsequent sections.

3.1. Path Loss. When sound propagates from underwater environment then some of its strength converts into heat. Sound wave propagation energy loss can be categorized into three main categories which are described below.

(1) Geometric Spreading Loss. When source generates acoustic signal it propagates away from the source in the form of wave fronts. It is independent of frequency, however, depending upon distance covered by wave front. Geometric spreading is divided into two types: first spherical spreading that depicts deep ocean communication; second cylindrical spreading that depicts shallow water communication [5, 18].

(2) Attenuation. Attenuation is defined as "wave energy converted into some other form of energy", such as heat energy, absorbed by the medium used. Within acoustic communication, this phenomenon is compassionated as acoustic energy is converted into heat. The converted heat is absorbed by underwater environment. Attenuation is directly proportional to frequency and distance [5, 18].

(3) Scattering Loss. Deviation regarding the line of sight of a signal or change in angle is generally a physical property. Underwater channel also contains this property that effects acoustic channel data transmission during communication. Surface roughness increases due to increase in the wind speed. That raises the end product of scattering surface. Scattering surface not only affects delays but also affects power loss [5, 18].

3.2. Noise. Noise can be defined as a quality of communication system that degrades signal strength of any communication system. In case of underwater acoustic channel there exist different kinds of noises. Underwater noises can be divided into two major categories. These are ambient noise and noises by human beings. Both kinds of noises are described in detail in the following sections.

(1) Noise by Human Beings. These noises are due to heavy machinery utilization, shipping activities, fishing activities, military activities, sonar activities, and aircraft activities and because of heavy data traffic sending and receiving activities cause different kind of disturbance and interference during acoustic communication. Sometime noises due to human beings also disturb natural acoustic communication [18].

(2) Ambient Noise. Ambient noise is a complex phenomenon regarding underwater communication. It can also be defined as a combination of different sources that cannot uniquely identify [22]. Ambient noise is also called background noise that occurs as a result of unidentified sources [5]. These noises are divided into four major categories which are known as wind, shipping, thermal, and the turbulence [1]. Wind noise is due to breakage of wave or because of bubbles created by air. Noise can be simply predicted and forecast from weather forecasts because of dependence of noise upon wind speed. Large number of ships present at large distance from communication system in ocean produce high traffic noise in acoustic communication, if sound propagation is good enough. Ships consider main source of anthropogenic ambient noise [22]. Turbulence can be defined as surface disturbance due to waves or tides that generates low frequencies that results continuous noise in acoustic communication. Underlying noise is considered as thermal noise in the absence of all other sources of noise, including self-noise. Thermal noise is directly proportional to the frequency which is used for acoustic communication [23].

3.3. Multipath. Sound propagation in shallow water is influenced by surface reflections while deep water propagation is affected by bottom reflection that becomes cause of large and variable communication delay in acoustic communication. A major cause that makes acoustic signal weak is called multipath effect that becomes cause of intersymbol interference which also makes acoustic data transmission difficult and erroneous. Vertical acoustic channel is less affected by multipath effect as compared to horizontal acoustic channel [18, 19, 21].

To address the problem of long propagation delay and high lite error rate a routing protocol QERP was proposed to handle end-to-end delay but this protocol still needs to address the mobility issues [14]. Mostly in deep oceans because of variable sound speed, refraction of sound occurs that cases of multipath effect in acoustic channel. Number

Table 2: Effect of temperature on acoustic communication.

S.No	Area Focused	Findings	Sound speed Effects due to temperature
01	Underwater Wireless Sensor Networks: Routing Issues and Future Challenges	Speed of sound increases due to increase in the temperature of ocean and decreases in colder oceans. Approximately, the mount of 1°c can boost the speed of sound near to 4.0 m/s.	Increase with temperature
02	Prospects and Problems of Wireless Communication for Underwater Sensor Networks	Shallow water effects acoustic communication by temperature gradients, ambient noises regarding surface and multi-path effect because of reflection and refraction.	Effects communication
03	Survey of temperature variation effect on underwater acoustic wireless transmission	Speed of sound is affected by temperature, depth and salinity of underwater environment. These factors produce variations in speed of sound in underwater environment.	Variation in speed
04	Variability of available capacity because of depth and temperature in the underwater acoustic communication channel	Determine acoustic channel capacity on short distances, increasing temperature and depth as a result gets higher channel capacity and throughput rates.	Improves throughput
05	Underwater Channel Simulation	Temperature of sea surface is much higher as is compared to the bottom temperature. Velocity of sound is also affected with increase in depth, salinity and temperature.	Increases with temperature
06	Mathematical equation for sound speed in the oceans	Temperature is a dominating factor that has effect on the sound speed.	Increases with temperature

of propagation paths, propagation delays, and its strength are determined by acoustic channel impulse response that is influenced by channel reflection and geometry. Large numbers of paths exist in acoustic channel but only those paths are considered which have less energy loss and reflections. All other paths are discarded as a result only finite number of paths remains for acoustic communication and data transfer [24].

3.4. Doppler Spread. Because of channel flaws, wireless signals practice a diversity of degradations. For example, electromagnetic signal affects by interference, reflections, and attenuation; acoustic signals regarding underwater are also affected by same kind of factors [25]. Underwater acoustic channel is complex channel due to time variation and space variation. The relative motion of transmitter and receiver that causes the mean frequency shift is called Doppler shift. Although the fluctuation of frequency in the region of this Doppler shift is called Doppler spread [8], two types of influences are observed on acoustic channel because of Doppler Effect: first is pulse width that will be compressed or stretched and second is frequency offset as a result of frequency offset compressing or expending of signal time domain occurring [26].

4. Related Studies

Presently underwater communication system utilizes electromagnetic, optics, and acoustic data transmission techniques to send data among different positions. Electromagnetic communication technique is affected by conducting nature of seawater while optic waves are applicable on very short distance because optic waves are absorbed by seawater. Acoustic communication is only one technique that has better performance regarding underwater communication due to less attenuation in seawater. Acoustic communication also has less attenuation in deep and thermally stable oceans. Shallow water affects acoustic communication by temperature gradients, ambient noises regarding surface, and multipath effect because of reflection and refraction [5]. Speed of sound is not constant in underwater environment instead of this speed of sound varies from point to point. Close to the surface of ocean speed of sound is found to be 1500 m/s that is four times higher than speed of sound in air but very slow as compared to speed of optics that is 3×10^8 m/s and electromagnetic in air. Table 2 shows the effects of temperature on acoustic communication and its effects with variation in temperature.

Natural acoustic systems and artificial acoustic systems both use acoustic channel in case of underwater environment. Both acoustic systems heavily utilize middle frequencies; because of that their communication affects each other, as they use same frequencies. Still, acoustic channel spectrum is not utilized efficiently. High spectrum utilization and to develop an environment friendly underwater acoustic network (UAN), Luo et al. [7] present Cognitive Acoustic (CA) as a promising technique. This technique has the capability to wisely sense whether any part of the spectrum is engaged by any other and also has the capability to change their frequency, power, or even other operation parameters to temporarily use the idle frequencies without interfering

with other networks. The CA technique makes communication environment friendly and erroneous free by avoiding interference with marine mammals. An important issue in underwater environment is use of low frequencies which results in low data rate. Other problems like energy dispersion and reflection also degrade the performance of devices. In this study the author proposed a model for underwater communication which monitors the performance of the wireless sensor nodes based on different frequencies and achieved high data rate [27].

Acoustic channel is highly variant because of unique challenges, e.g., narrow bandwidth, long propagation delays, variable speed of sound, reflection, refraction, and large propagation losses. These unique challenges also create problems regarding media access control protocols. Media Access Control protocols have two main categories these are scheduled protocols and contention-based protocols. Scheduled protocols avoid collision among transmission nodes, while in contention-based protocols nodes compete each other for sharing a single channel. Scheduled based protocols, for example, Time Division Multiple Access (TDMA), are not efficient due to large propagation delays; frequency division multiple access (FDMA) is not suitable due to the narrow bandwidth; and Code Division Multiple Access (CDMA) is suitable for underwater acoustic networks. While contention-based protocols are not appropriate for underwater communications [9], Lv et al. [28] propose TDMA based Underwater Acoustic Channel Access Control method (UA-MAC), to improve channel utilization in dense Mobile Underwater Wireless Sensor Networks (MUWSN). Aim is to solve the difficulties like, time schedule to access the channel, hidden terminal problem, and end-to-end delay. Underwater acoustic channel access method puts into practice the piggyback scheme and as a result fewer packets are exchanged. Using that kind of methodology, collision decreases and saves a lot of energy. Shahab-u-deen et al. [29] combine different media access control protocols in a suite called Adaptive Multimode Medium Access Control for Underwater Acoustic Networks, because no single protocol can complete the requirements of underwater sensor network media access control. Adaptive Multimode Medium Access Control for Underwater Acoustic Networks aims to improve the performance regarding traffic intensity. This suite switches from one protocol to the other based on network requirements, traffic intensity, and quality-of-service requirements.

Channel capacity is affected by temperature, depth, propagation loss, and ambient noise of underwater environment where sensor nodes are deployed. Path loss is the function of distance (between pair of nodes) and frequency utilized for communication. These factors affect acoustic channel capacity. However, bandwidth increases with increase in depth and temperature and decreases with increase in distance. Sehgal et al. [30] determine acoustic channel capacity on short distances, increasing temperature and depth as a result gets higher channel capacity and throughput rates. Large bit error rates and large delays are characteristics of acoustic channel. Harris et al. [31] compare three different techniques adaptation of packet size, forward error correction, and adaptation of packet train size to overcome long delays and large bit error rates and also to improve channel utilization. Packet train length overcomes long propagation delays in addition of time wastage while packet size adaptation and forward error correction overcome both large propagation delays and bit error rates. Acoustic channel utilization also increases under the utilization of packet size adaptation and forward error correction. Harris's analysis provides guidelines for creation of media access control and routing protocols.

Information regarding sensor nodes is useful only when localization is involved in it. Large numbers of terrestrial localization schemes are available but because of unique challenges (sensor nodes movement with ocean currents, high cost of senor nodes, global position system inapplicability, and limited battery power) of underwater sensor networks they cannot be utilized directly. Guo et al. [32] provide a mechanism of localization which is known as Anchor-Free Localization Algorithm (AFLA). This algorithm has ability of self-localization for anchor-free sensor nodes. AFLA uses anchor nodes and cables to restrict sensor node in underwater environment. AFLA's goal is to create an efficient localization scheme for underwater sensor networks. Simulation results prove that AFLA is an efficient localization scheme and it can be utilizable in both static and dynamic networks scenarios. Table 3 highlights the major effects of noise and bit error rate during acoustic communication using different protocols.

Major issues, e.g., energy conservation and mobility regarding underwater sensor networks, create unique challenges for designing of routing protocols and make all existing ground-based routing protocols (proactive and reactive) inadequate. Underwater environment required such protocols that are efficient in energy consumption, manage random variation in topology, and consider asymmetric links and huge propagation delay. DU et al. [33] present a protocol which is known as Level-Based Adaptive Geo-Routing (LB-AGR) that divides communication traffic into four categories. These are upstream to sink, downstream to sensor nodes, downstream to specific nodes, and downstream to all nodes. Data forwarding is based upon density, available battery power, and level between neighbors that is used to elect next best hop. Level-Based Adaptive Geo-Routing goal is to achieve minimum communication delay, consume less battery power, and improve delivery ratio as well as received packets percentage. This protocol reduces communication end-to-end delays and improves delivery ratio and efficient utilization of battery power. Efficient utilization of battery power is the major concern of underwater sensor networks routing protocols.

Huang et al. [34] proposed a routing protocol that utilized energy efficiently using fuzzy logic and decision tree techniques for data forwarding towards the surface sink. Routing protocol goal is to utilize battery power efficiently in that manner that reduces the expenditures of energy during acoustic communication. Protocol reduces traffic overload on acoustic channel and reduces energy consumption also. Presently, for routing protocols minimum end-to-end delay and high efficiency are the major requirements for underwater sensor networks. Ali et al. [35] present an end-to-end delay

Table 3: Effect of noise, errors, and protocols.

S. No.	Area of research	Findings	Noise effect	Bit error rate	Protocol usage
01	Challenges: Building Scalable and Distributed Underwater Wireless Sensor Networks (UWSNs) for Aquatic Applications,	Time Division Multiple Access (TDMA) is not efficient due to large propagation delays, Frequency Division Multiple Access (FDMA) is not suitable due to the narrow bandwidth and Code Division Multiple Access (CDMA) is suitable for underwater acoustic networks. While contention-based protocols are not appropriate for underwater communications.	N/A	N/A	(TDMA) is not efficient; (FDMA) is not suitable. (CDMA) is suitable
02	Prospects and Problems of Wireless Communication for Underwater Sensor Networks	Acoustic communication also has less attenuation in deep and thermally stable oceans. Shallow water effects acoustic communication by temperature gradients, ambient noises regarding surface and multi-path effect because of reflection and refraction.	Decreases by ambient noises	less attenuation in deep and thermally stable oceans	N/A
03	Analyzing the Performance of Channel in Underwater Wireless Sensor Networks (UWSN)	Multi-path propagation results in fading and phase fluctuations, Doppler Effect is observed due to the movement of both the sender and receiver nods. Speed of sound and underwater noise are other factors that influences the performance of acoustic channel.	Decreases by ambient noises	Fading and phase fluctuations, Doppler Effect	N/A
04	Optimized packet size selection in underwater wireless sensor network communications	Effect of bit error rate, interference, collision, retransmission leading selection of optimal packet size is also considered and achieves improvement in all metric e.g., throughput, energy consumption, resource utilization and packet latency underutilization of optimal packet size selection.	N/A	Less bit errors in small packets	N/A
05	Choosing the packet size in multi-hop underwater networks	Data packets are large enough as compared to the control packets and because of control and data packet collision entire data packet is discarded that causes of huge number of retransmissions and energy dissipation.	N/A	Error due to control and data packet collision entire data packet discarded	CDMA, DACAP
06	Challenges for efficient communication in underwater acoustic sensor networks	A major cause that makes acoustic signal weak is called multi-path effect, that becomes cause of inter-symbol interference also makes acoustic data transmission difficult and erroneous. Vertical acoustic channel is less affected by multi-path effect as compared to horizontal acoustic channel.	N/A	Interference and multi-path effect	N/A
07	Ocean ambient noise: Its measurement and its significance to marine animals	Large number of ships present at large distance from communication system in ocean produce high traffic noise in acoustic communication, if sound propagation is good enough. Ships consider main source of anthropogenic ambient noise.	N/A	Decreases due to ships noise	N/A
08	SEA 6 Technical report: Underwater ambient noise	Turbulence can be defined as surface disturbance due to waves or tides that generates low frequencies that results continuous noise in acoustic communication. Underlying noise is considered as thermal noise in the absence of all other sources of noise, including self-noise. Thermal noise is directly proportional to the frequency which is used for acoustic communication.	surface disturbance of waves or tides generates low frequencies that results in noise in acoustic communication	Errors due to noise	N/A
09	Doppler estimation and correction for shallow underwater acoustic communications	Because of channel flaws, Wireless signals practice a diversity of degradations. For example, electromagnetic signal affects by interference, reflections, and attenuation, acoustic signals regarding underwater are also affected by same kind of factors.	N/A	High BER due to interference, reflections, and attenuation	N/A

Table 3: Continued.

S. No.	Area of research	Findings	Noise effect	Bit error rate	Protocol usage
10	Study on Doppler effects estimate in underwater acoustic communication	Two types of influences are observed on acoustic channel because of Doppler Effect, first is pulse width that will be compressed or stretched and second is frequency offset as a result of frequency offset compressing or expending of signal time domain occurs.	N/A	Doppler Effect, frequency offset	N/A

efficient routing protocol which is known as Diagonal and Vertical Routing Protocol for Underwater Sensor Network (DVRP). Packet forwarding mechanism is depending upon angle of flooding zone and flooding nodes are also controlled by manipulating the angle for flood region to avoid the flooding over the entire network. Diagonal and Vertical Routing Protocol goal is to minimize end-to-end delay and consume less battery power of sensor nodes. Diagonal and Vertical Routing Protocol has no need to maintain large routing tables; instead of this it uses its local information to rout data packet towards destination. Adding or removing new nodes create no disturbance for existing nodes.

Basagni et al. [12] select two protocols. These are Code Division Multiple Access (CDMA) and Distance Aware Collision Avoidance Protocol (DACAP) and compare them by varying packet size, bit error rate, and traffic load. Under the exploitation of variant packet sizes, bit error rates, and traffic load, author determines the impact of packet size upon multihop underwater sensor networks. Basagni et al. conducted an experiment using different fixed (predetermine) packet sizes and packet size is not change according to environmental factors. Data packets are large enough as compared to the control packets and because of control and data packet collision entire data packet is discarded that causes of huge number of retransmissions and energy dissipation.

Basagni et al. [11] introduce technique of data fragmentation that minimizes disadvantage of collision by partitioned long data packet into small fragments. Technique is experimented upon DACAP by considering and nonconsidering data fragmentations. Fragmentation decreases retransmissions, energy consumption, and packet latency as well as overall traffic and huge overhead. Basagni et al. do not consider the factor of varying bandwidth and interference and as a result of fragmentation traffic load increases that congested communication channel. Since improving the throughput of a single hop packet size is considering a critical parameter regarding communication in field of underwater sensor networks. Underwater sensor networks use the half duplex methodology for communication that can be avoided by utilizing optimal packet size selection. In this proposed feeding control system sensor nodes works in groups for necessary decision making. The contribution of this system is to avoid loss of food and then reduces negative impact on environment as well as economically feasible [27, 36].

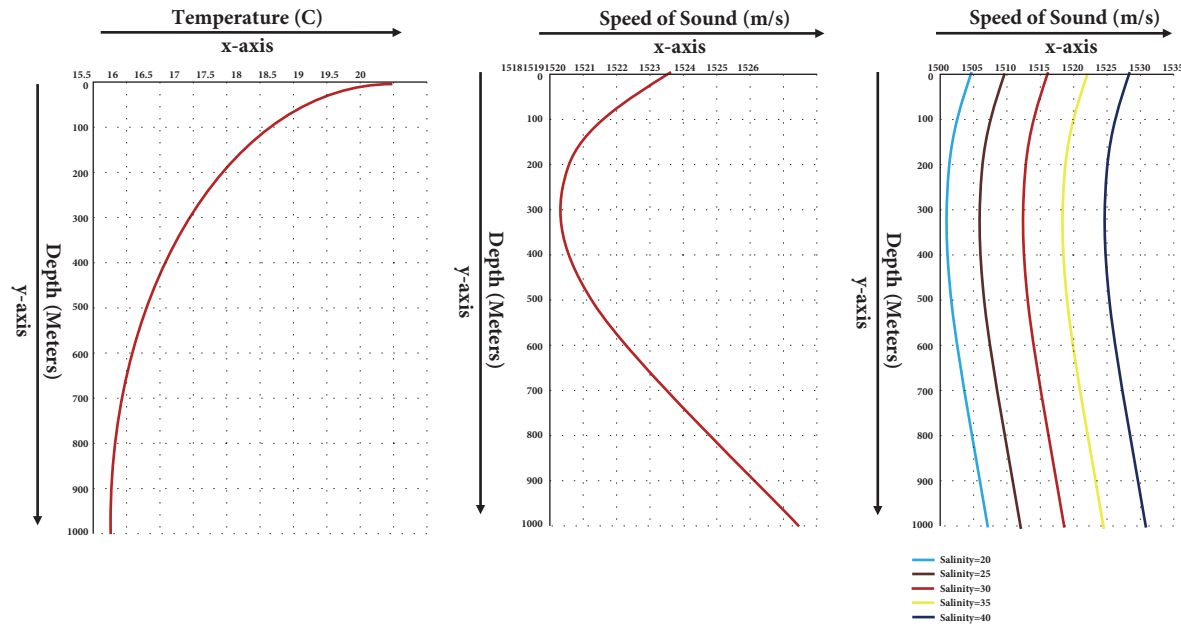
QERP is proposed by Faheem et al. [15] to improve the reliability of data transfer in underwater acoustic sensor networks. The mechanism used for organizing sensor nodes is in the form of small clusters which are connected hierarchically for distributed energy and data transfer evenly. This technique reduces the probability of packet loss and preserves high link quality in underwater environment. The issue with this technique is no mobility and node density is not addressed [37].

Basagni et al. [10] observe the performance of multihop network in provisions of throughput, energy efficiency, and latency. Effect of bit error rate, interference, collision, and retransmission leading selection of optimal packet size is also considered. For this, Basagni et al. select two-media access control protocol CDMA and DACAP and compare their results and change the network deployment scenario and then observe the effect of the packet size upon throughput, energy efficiency, and latency of network. Basagni et al. achieve improvement in all metric, e.g., throughput, energy consumption, resource utilization, and packet latency underutilization of optimal packet size selection. Junget al. [38] have investigated energy efficiency using optimal packet size under the utilization of NS-2 simulator. Authors create a cluster of 100 nodes in dimensions of 2 km×2 km×200 m. Experiment proves a relationship between energy efficiency and packet size. Optimal packet size reduces utilization of extra energy. Erroneous channel offers large bit error rate that causes wastage of large energy amount, but the utilization of optimal packet size for erroneous channel reduces wastage of energy. Reliability (in sense of data delivery) is a major issue regarding underwater sensor networks because of highly variant environment.

Ayaz et al. [13] provide an algorithm that has the ability to determine the best suitable packet size for reliable data transfer, using two-hop acknowledgment methodology for same packet size. Algorithm investigates optimal data packet size for underwater sensor networks, with energy efficiency as the optimization metric. The goal of the algorithm is reliable data delivery from source node to destination node or surface sink is a major requirement of a network. In this section, paper presents a brief overview of different advancements in the field of Underwater Wireless Sensor Networks that uses acoustic channel for communication. But still some issues and challenges exist. That not only affects the performance of above described methodologies but also needs solution from research community. These issues, challenges, and drawbacks are discussed in the next section.

5. Issues and Challenges regarding Underwater Sensor Networks

In this section, issues and challenges regarding underwater sensor networks are described which makes underwater communication hard and problematic as compared to



(a) Temperature versus depth (b) Depth versus sound velocity (c) Salinity versus sound velocity

Figure 3: Variation of environmental factors (depth, temperature, and salinity).

terrestrial sensor network communication. In underwater sensor network issues and challenges like reliability and efficient utilization of acoustic communication link, optimal packet size selection for communication, power consumption, distributed localization, environmental effects, media access control, and network routing protocol still need to address. These problems require attention from academia and research community. Following section describes efforts which are made by researchers to address these issues.

5.1. Environmental Effect. Marine organisms are affected due to anthropogenic sound which is emitted in underwater environment, in various ways; e.g., organs of hearing are affected in shape of hearing loss; high potential sound waves which are received by marines can injure and also can become the cause of their death [52]. Presently, underwater communication utilizes different kind of communication methodologies, e.g., optic, electromagnetic, and acoustic. Only acoustic communication meets the requirements of underwater communication due to less attenuation and low absorption in sea water [5].

Speed of sound is not constant in ocean. Near to the surface of the ocean the speed of sound is 1500 m/s that is four times, as compared to the sound speed in air. Upper part of ocean is called surface layer, in which change of temperature is less significant while in layer beneath (thermocline) temperature is prominent factor which affects the speed of the sound as compared to others [53].

Speed of sound increases due to increase in the temperature of ocean and decreases in colder oceans. Approximately, the mount of 1°c can boost the speed of sound near to 4.0 m/s. The boost of 1 practical salinity unit can enhance the speed of sound nearly 1.4 m/s; when we walk off deep, the pressure of ocean water continues to enhance; as a result each 1 km depth will boost the sound speed of nearly 17 m/s [4]. In this survey authors discussed issues in underwater environment and observed that acoustic waves are used to get different environment parameters and there is no standardization of parameters for the monitoring of underwater environment, which results in the variety of monitoring system but none of them is a standard [54].

Temperature of sea surface is much higher as is compared to the bottom temperature. As is shown Figure 3, temperature is falling with depth and then becomes constant. Velocity of sound is also affected with increase in depth, salinity, and temperature, because these are the major factors that affect speed of sound in underwater environment. The effect of these environmental parameters can be observed in three domains. In first domain the impact of temperature is dominating, as compared to the other parameters, but in second domain both depth and temperature are dominating factors upon the sound speed. In third domain, the sound speed is purely dominated by the depth. Speed of sound is also depending upon salinity. Speed of the sound increases with the increasing salinity of the sea water, but the shape of the curve does not change [55, 56].

Underwater communication networks must utilize such kind of sound waves which do not affect natural acoustic communication and the organs of the water creature. In the development of underwater communication technique utmost care must be taken regarding the life of marine animals (organ) and their communication. It is still an open area for research and it needs solution from research community.

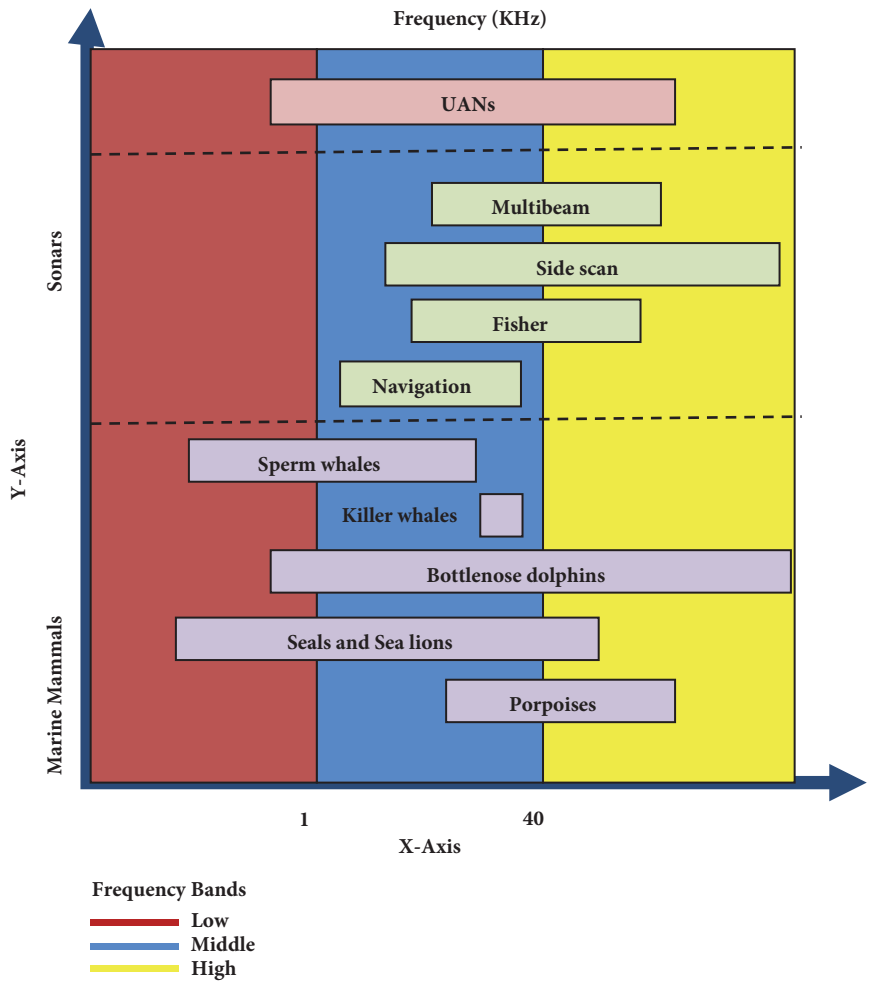


Figure 4: The acoustic spectrum usage in underwater environment.

5.2. Cognitive Communication. Because of attenuation, the available communication frequencies are mostly limited, usually from tens of hertz to hundreds of kilohertz in underwater environment. Mostly artificial acoustic systems and natural acoustic systems utilize the frequency band from 1 kHz to 100 kHz that makes the acoustic channel crowded. Frequency utilized by artificial acoustic systems is overlapping with natural acoustic systems, for example, marine mammals.

The underwater acoustic system overlapped spectrum usage is summarized in Figure 4. That indicating the fact about the underwater spectrum, especially on mid-frequencies, is heavily shared by various acoustic users. Still, acoustic spectrum is temporally and spatially underutilized in underwater environments.

Unfortunately, existing underwater acoustic networks designs have mainly focused on the single network scenario, and very few have considered the presence of nearby acoustic activities. Furthermore, the mobility and idle listening of acoustic systems may cause the spectrum to be underutilized temporally. The directional transmission and reception of acoustic systems may potentially cause the spectrum to be underutilized spatially. For high spectrum utilization and to develop an environment friendly with underwater acoustic networks, CA is presented as a promising technique. This technique can intelligently detect whether any part of the spectrum is engaged by any other. Also, it can change its frequency, power, or even its other operational parameters to temporarily use the idle frequencies without interfering into other networks.

The CA technique makes communication environment friendly by avoiding interference with marine mammals. Hence, each user first senses the surrounding spectrum before sending and receiving in underwater acoustic networks. Sending and receiving only take place when the sensing frequencies are idle. Thus, CA users can prevent using frequencies, which are engaged by marine mammals and look for new idle frequencies for communications. When marine organisms discharge the spectrum, CA users reuse these idle frequencies again. There exist some recently developed protocols which support dynamic spectrum management while underwater CA network (UCAN) is still underexplored area. Presently, CA technique is suffering from different features of underwater environment, e.g., long propagation delay, narrowband, long preamble embedded by acoustic modems,

severe busy terminal problem of acoustic modems, and highly dynamic underwater channel. The unique characteristics of underwater channel provide challenges for the design of CA network. These challenges are spectrum sensing, dynamic power control, spectrum sensing strategy, spectrum sharing, and spectrum decision [7].

Efficient utilization of acoustic spectrum is a major requirement of the time because partial utilization makes acoustic communication very congested upon central frequencies that is mostly utilized by natural acoustic systems. Future developments regarding acoustic channel which will be aware of acoustic spectrum will make better utilization of unutilized parts of acoustic spectrum.

5.3. MAC Issues. Acoustic channel affects from various problems, for example, narrow bandwidth and huge transmission delays. Furthermore, acoustic channel has exclusive challenges regarding media access control. Media Access Control protocols have two main categories. First is called scheduled protocol that has avoided collision between transmission nodes. Second is known as contention-based protocols where nodes compete each other for sharing a single channel.

Working of different protocols varies with respect to environment. FDMA is not appropriate because of the constricted bandwidth of acoustic channel for underwater, whereas TDMA is not efficient, due to large propagation delays of acoustic channel. Several antenna elements are deployed at convinced access points, then Spatial Division Multiple Access (SDMA) is also an applicable option [9]. Underwater sensor networks are characterized as containing long propagation delay and low data rate. TDMA based underwater acoustic channel access method (UA-MAC) is proposed to improve channel utilization in dense Mobile Underwater Wireless Sensor Networks (MUWSN). According to the analysis of Lv et al. and depending upon the challenges of media access control protocol in MUWSN, underwater acoustic channel access method aims to solve the difficulties like, time schedule to access the channel, hidden terminal problem, and end-to-end delay. Moreover, underwater acoustic channel access method which is implementing piggyback method and fewer packets are exchanged, because of this reason, collision decreases, and a lot of energy is also saved. Reliable data delivery is the major concern for data gathering, navigation, and observation of the ocean regarding the field of underwater sensor networks [28].

5.4. Channel Utilization. Designing a highly utilizable channel is a great challenge, due to the characteristics of underwater environment, for example, multipath propagation which results in fading and phase fluctuations. Doppler Effect is another problem which is observed due to the movement of both the sender and receiver nods. Speed of sound and underwater noise are, further, factors which influence the performance of acoustic channel [8].

Sehgal et al. [30] analyzed the channel capacity which depends upon depth and temperature by considering propagation loss and ambient noise. Sehgal et al. compare attenuation modes as Thorp's model with Fisher and Simmons model. Thorp model considers fixed values of depth and temperature and not considering any change regarding depth and temperature, for the calculation of acoustic channel capacity. However, on the other side Fisher and Simmons model considers change regarding depth and temperature. Results show that channel capacity and utilization of acoustic channel increases with high throughput because of increase in both temperature and depth in case of short distances. Sehgal et al. also describe that nodes in deep ocean get high bandwidth and higher throughput rates. Higher channel capacity utilizes the same power with increasing depth. Bandwidth and channel capacity depend upon temperature and depth. Networks those have small area to cover get benefit from this model. Although for the networks which cover long distances no solution is provided in this sense. Same power is consumed for higher bandwidth capacity.

Due to bandwidth dependency upon the transmission distance, we get huge throughput if messages are forwarding using multihops instead of transmitting straight forwardly using one long single hop. During the analysis of [57], regarding acoustic channel physical model, under the consideration of propagation and ambient loss, Stojanovic determines bandwidth dependency upon distance. Underwater acoustic channels have the characteristic of path loss that depends not only upon the space among the nodes, but also upon frequency of the signal. Signal frequency can be utilized to measure the absorption loss of the communication signal, which occurs because of conversion of acoustic power into warm-up. Short communication link provides more bandwidth as compared to a longer one in an underwater acoustic system. Hence, utilizing this technique we get significant increase in information throughput and acoustic channel utilization efficiently. Simultaneously, less energy will be consumed; however, that phenomenon is also true regarding the radio channel.

Analysis of Harris et al. [31] depicts that use of three methods, adaptation of packet size, forward error correction, and adaptation of packet train size, belongs to improvement regarding acoustic channel utilization for underwater environment. MAC and routing protocols must consider energy consumption as well as end-to-end delays. Simulation's results have proved that utilization of these techniques raises increase in channel utilization in the presence of underwater acoustic channel constraints. First technique overcomes long delays of acoustic channel; however, for this effectiveness, network must compensate in the shape of time. Time exceeds prior to the sender can be reported of a failure. Alternatively, forward error correction and adaptation of packet size are inducting to overcome both the propagation delay and high bit error rates. Finally, we conclude these three kinds of methodologies; optimal selection of parameters depends upon the distance among the source and destination nodes, which affects both the propagation delay and the bandwidth for the underwater acoustic links utilizable channel is a great challenge, due to the multipath propagation, Doppler Effect, underwater noise, and dependence of channel capacity upon distance. In the presence of all these issues and challenges regarding underwater environment, acoustic communication

Table 4: Performance evaluation of LB-AGR with existing protocols of UWSNs.

Factors	LB-AGR	VBF	VBVA
Energy Consumption	Low	High	High
Successful packet received	High	Low	Low
End-to-End interruption	Low	High	High
Delivery percentage	High	Low	Low

demands methodologies from research community that better utilize the acoustic channel [58, 59].

5.5. Localization. In underwater acoustic networks, localization of sensor nods is not required because they are fixed by utilizing anchors or tied with surfaces buoys in the assistance of Global Positioning System (GPS). Although on the other side for underwater sensor networks, major problem is localization, mostly sensor nods are movable by means of the current of the ocean. Shaping accurately portable sensors nodes, locations in oceanic situations are more demanding as compared to underwater acoustic networks [9]. However, localization is the major issue in underwater sensor network scenario, because GPS signals (electromagnetic) do not efficiently propagate through sea water. Only acoustic communication is feasible in underwater environment. Localization techniques for underwater sensor networks can be categorized as range based and range free techniques, static reference nodes and dynamic reference nodes, and single stage and multistage schemes. All techniques perform well in simulations but they are not experimented in same conditions or assumptions [60]. This review article discussed the state-of-the-art localization based and localization free routing protocol. The major problem in routing is limited bandwidth, energy consumption, propagation delay, and short of memory [61].

Existing schemes consider a predefined locations map, using anchored nodes, for controlling the locations of sensor nodes. In deep water scenarios, special kind of nodes which get commands from control stations, for example, AUVs or surfaces buoys, is used. All these mechanisms are very costly, and their performance is not much efficient. Alternatively, in terrestrial sensor wireless networks every node maintains its location later than deployment. In contrary, sensor nodes in underwater scenarios do not maintain their location after deployment. Instead of this they shift by means of ocean current, tide, and other aspects. To regulate the shun nodes from deviation, from the place of deployment, generally, attach them to predetermined anchors through cables.

Thus, the movement of underwater nodes is actively restricted. That motivates researches and they provide an idea of Anchor-Free Localization Algorithm that is called AFLA. AFLA is considered for sensor networks which are actively restricted in underwater environment. Anchor node's information does not require by AFLA, and constructs employ of the association of neighboring nodes. In both static and dynamic network scenarios AFLA can be utilized. This algorithm contains a self-localization mechanism for underwater anchor-free sensor nodes. It can localize all nodes without anchor node's assisting [32]. Although, this algorithm has efficient results in underwater scenario but the localization of a freely moving node is still an open area for research. Data is only meaningful when exact location information is attached with it.

5.6. Routing Issues. A major issue that is affecting underwater sensor networks is energy saving. Nodes mobility is another challenge in underwater sensor networks. Different protocols regarding land-based sensor networks are, for example, Directed Diffusion, Gradient, Rumor routing, TTDD, and SPIN. These protocols (Directed Diffusion, Gradient, Rumor routing, TTDD, and SPIN) are generally planned for at a standstill network. However, due to the mobility and very rapid change in "network topology" make, these existing ground-based routing protocols are insufficient for underwater environment [9].

Offered routing protocols for earthly sensor networks are separated into two extensive groups, Proactive and Reactive. Except, both have problems in underwater environment. Table Driven or Proactive protocols construct huge signaling overhead to set up the routs, more than ever for the first and each time when the topology is customized. Topologies are modified always, by the reason of continuous nodes movement, and due to the large acoustic signaling delays proactive protocols are not efficient solution for underwater environment. Because of huge acoustic delays and asymmetrical links, consequently, the protocols of these (Reactive scheme) natures are not appropriate for the underwater networks [4, 62]. Researchers make efforts to design such protocols which can perform better in underwater environment.

DU et al. [33] showed that Level-Based Adaptive Geo-Routing (LB-AGR) protocol traffic is divided into four categories. These categories are upstream to the sink, downstream to sensor nodes in elected area, downstream to specific node not considering of where the node is locating, and downstream to all the sensor nodes. Upstream of packets towards sink is forwarded in unicast manner to the best next hop, instead of broadcasting to every neighbor node as compared to present underwater sensor network routing protocols (VBF and VBVA). Level-Based Adaptive Geo-Routing considers the following factors for forwarding which are available energy, density, location, and level-difference between neighbor nodes which are used to determine the best next hop between multiple qualified candidates. Performance evaluations of Level-Based Adaptive Geo-Routing with existing protocols of underwater sensor networks are given in the Table 4.

Table 5: Comparison of packet fragmentation versus nonfragmentation.

Factors	Fragmentation	Non-Fragmentation
Delay	Low	High
Energy Consumption	Low	High
Traffic load	High	Low
Channel utilization	Low	High
Collision effect	Low	High

Forward tree trimming methodology is implemented, to avoid overload increase by forwarded packets and also to efficiently shrink the voltage expenditure of the sensor nodes [34]. A protocol that provides higher level of efficiency in the environment that is highly volatile, for example, underwater environment, is a major requirement of the time. An end-to-end delay efficient routing protocol which is known as Diagonal and Vertical Routing Protocol (DVRP) for underwater sensor network is presented. Data packet forwarding is depending upon the angle of flooding zone via the sender node, in the direction of the surface sink, in case of DVRP. Alternatively, the quantity of nodes that floods the packets of data is prohibited by manipulating the angle for flood region to avoid the flooding over the entire network. For increasing the life time of the network, DVRP is consciously preserving sensor nodes energy. In this (DVRP) network sensor nodes make local judgment of forwarding data packets underneath the restriction of energy status and flooding angle among them. Protocol does not keep the multifaceted routing tables and not depend on information regarding location. It is easy to include a fresh node, at any instance and any position in the network [35].

Routing is a primary concern for any category of network, and routing protocols are utilized for maintaining and discovering the paths. Underwater sensors network contains the issue regarding underwater network physical and network layer [62]. However, routing techniques regarding network layer are a new research area.

5.7. Optimal Packet Size Selection and Energy Efficiency. Basagni et al. [12] explore the packet size impact upon two protocols and measure end-to-end delay; throughput efficiency and energy per bit are utilized. Optimal packet size is depending upon protocol characteristic, just like offered load and bit error rate. Poor packet size selection decreases the performance of the network. Recently, the emphasis has shifted towards the multihop network, because it covers large area as compared to single hop networks. Optimal packet size selection is considered to increase the efficiency of that kind of networks. Stojanovic (upon point-to-point scenario at data link layer) has explored packet length optimization for maximizing throughput efficiency. Results have proved that packet size affects the performance of network.

Metrics such as throughput efficiency, latency, and energy consumption in multihop underwater networks can be greatly improved by a wisely choosing packet size. Basagni et al. use variable but predefined fixed size of the packets and determines the performance of CDMA and DACAP. Basagni et al., not changing the packet size point-to-point, means that, according to environmental factors and channel capacity at different parts of the ocean, using this kind of technique, performance and channel utilization will also increase. Basagni et al. [11] showed that size of control packets is very small as compared to the data packets. Once a data and control packet collision occur, the entire data packet must be discarded. Long data packet can be partitioned into smaller fragments, to minimize disadvantages of such kind of collisions. Effect of collision is reduced and remains to only few fragments by utilizing fragmentations; in that manner small numbers of chunks need to be retransmitted. Fragmentation reduces the overall traffic, as well as higher overhead and the number of retransmissions as well. This technique is applied upon DACAP. Advantages are achieved by fragmenting long packets, except because of fragmentation overhead increases in the network. For this reason, there exist an optimal number of fragments considering throughput efficiency. As traffic load increases, collisions also become higher.

Collision reduces because of fragmentation; another advantage of fragmentation is that overall energy consumption also reduces. Conversely, higher overhead is introduced for a single data transmission. Throughput efficiency decreases due to high traffic and higher bit error rate. However, the energy waste is considerably less as compared to no fragmentation. Packet fragmentation also reduces packet latency as shown in Table 5. Because of choosing optimum number of fragments higher throughput efficiency, lower energy consumption, and shorter packet latency are achieved.

Underwater sensor networks stop and wait mechanism, which is utilized in half duplex mode, can be avoided by using optimized packet size selection. Optimized packet size selection has positive impact upon multiple hope communication.

Basagni et al. [10] observe the performance of multihop network in terms of throughput, energy efficiency, and latency. They also considered the effect of bit error rate, interference, collision, and retransmission upon selecting the optimal packet size. Authors in [10] selected two-media access control protocol CDMA and DACAP and compares their results. In addition, variation in network deployment scenarios is observed to check the effect of the packet size upon throughput, energy efficiency, and latency of network. However, three fixed bandwidths are considered in all simulations despite the fact of point-to-point change in underwater bandwidth. The use of variable bandwidths is just like the use of variable data payloads by Basagni et al. [10], for better performance.

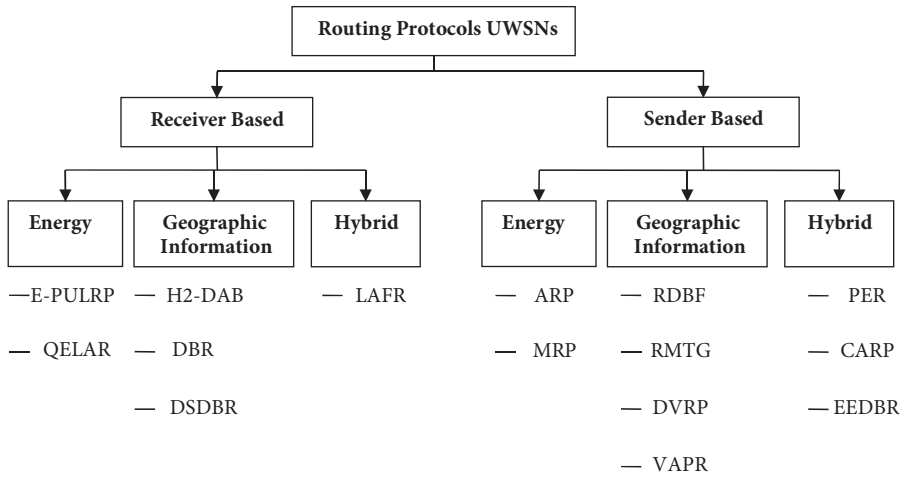


Figure 5: Classification of the routing protocols for UWSN.

Jung et al. [38] describe issues of energy efficiency and optimal packet size selection for efficient communication in underwater sensor networks. Challenges in underwater sensor networks are fading, multipath propagation delay, limited bandwidth, and energy constraints. Packet loss rate is higher because of acoustic channel refraction, reflection, and ambient noise. Ayaz et al. [13] have investigated optimal data packet size for underwater sensor networks, with energy efficiency as the optimization metric. Major challenges in Underwater Wireless Sensor Networks are high channel impairments, due to which higher error rates and temporary path losses occur in underwater sensor network. Congestion increases upon the nodes near the sinks, as data packets are forwarded towards the surface control station. Packet loss is caused by congestion which requires retransmissions that causes the loss of a significant amount of energy and leads to large end-to-end delays.

Packet size selection according to variations of environmental factors also improves acoustic communication and needs attention from research community. The communication techniques, which consider variations of underwater environment regarding acoustic channel, will produce better effect as compared to those which never consider these effects highly utilizable acoustic channel which is a great challenge, due to the multipath propagation, underwater noise, dependence of channel capacity upon distance, and Doppler Effect. Due to the movement of sensor nodes under the influence of ocean currents, meaningful information collection is also an issue that is called localization issue. Localization issue and frequent change in topology produce unique challenges for routing protocols. Packet size selection is a critical parameter regarding underwater sensor networks that can enhance energy consumption, channel utilization, and throughput of underwater sensor network. Hence these are the unique issues and challenges that need attention from research community. Routing protocols for UWSNS are classified in Figure 5.

Table 6 [63] shows that the efficiency in transmitting the packets is measured by publication year, bounded latency, multipath, load balancing, and energy consumption. In E-PULRP [39] and MRP [46] duty cycles are adopted for efficient energy consumption, while in QELAR [40] and LAFR [44] routes having minimum energy cost are selected by sensor nodes for purpose of energy transmission. Efficient energy consumption is the main objective for designing the UWSN protocols. Few of protocols discussed in Table 6 consider the load balancing technique to prolong the network life time. For data centric networks in UWSNs sensed data must be accurately transferred in bounded latency to the sink due to higher propagation delays in acoustic signals. Transmission delays are carefully analyzed in ARP [45] in which various priorities are assigned for packet transmission.

Multipath routing like DBR [42], MRP [46], and EEDBR [43] performs far better compared to single path routing protocols like QELAR [64] and H2-DAB [65].

Investigation of MAC protocol is shown in Table 7. Different MAC protocols which compared some of them are contention-based protocols like T-Lohi., R-MAC, MACUWASN, and MAFAMA and other or hybrid of contention based and contention-free like H-MAC, UWAN-MAC, and P-MAC.

In Figure 6(a), it is shown that DBR are only depth-based routing protocol and while calculated costs considering different routing metrics are ERP2R and EEF. Common metrics for the all mentioned protocols is depth. Overhearing is used in ERP2R to make transmission decision during holding time beside this in EEF Overhearing is avoided due to which the efficiency of EEF is uplifted.

Total energy consumption is shown in Figure 6(a) for three contention avoidance MAC protocols, CSMA-CA, MACU WSN, and T-LOHI, for different packets numbers. Similar increasing energy consumption trend has been seen in each protocol. MACU WSN has less control packets as compared to CSMA-CA and T-LOHI.

In Figure 6(b) comparison of VBF, REE-VBF, and HH-VBF has been shown. Vector based routing scheme is followed by these routing protocols; per hop vector is seen in HH- VBF that decreases the number of nodes involved in

Table 6: Comparison of the underwater routing protocols.

Routing Protocol	Publication Year	Bounded latency	Multi-path	Load balancing	Based		Energy Consumption	Geographic information		Communication overhead	Time complexity
					Receiver	Sender		Location	Depth		
E-PULRP[39]	2010	×	×	×	√	×	Medium	×	×	Low	O(n)
QELAR [40]	2011	×	×	×	√	×	Low	×	×	Low	O(m1)
H2-DAB [41]	2009	×	×	×	√	×	Low	×	√	Low	O(n2)
DBR [42]	2008	×	√	×	√	×	High	×	√	Low	O(m1)
DSDBR [43]	2014	√	√	×	√	×	Medium	×	√	Low	O(m1)
LAFR [44]	2013	×	√	×	√	×	High	×	×	Low	O(n×m1)
ARP [45]	2008	√	√	×	×	√	High	√	×	High	O(m1)
MRP [46]	2013	×	√	×	×	√	Low	×	×	Low	O(m2)
RDBF [47]	2014	×	√	√	×	√	High	√	×	High	O(n×m1)
RMTG [48]	2011	×	√	×	×	√	High	√	×	High	O(n2)
DVRP [35]	2014	×	×	×	×	√	High	√	×	Median	O(m1)
VAPR [49]	2013	×	√	×	×	√	High	×	√	Median	O(m1)
PER [34]	2011	×	√	×	×	√	Medium	×	×	High	O(n2)
CARP [50]	2015	×	×	×	×	√	High	×	×	High	O(m1)
EEDBR [51]	2012	×	√	×	×	√	Medium	×	×	High	O(m1)

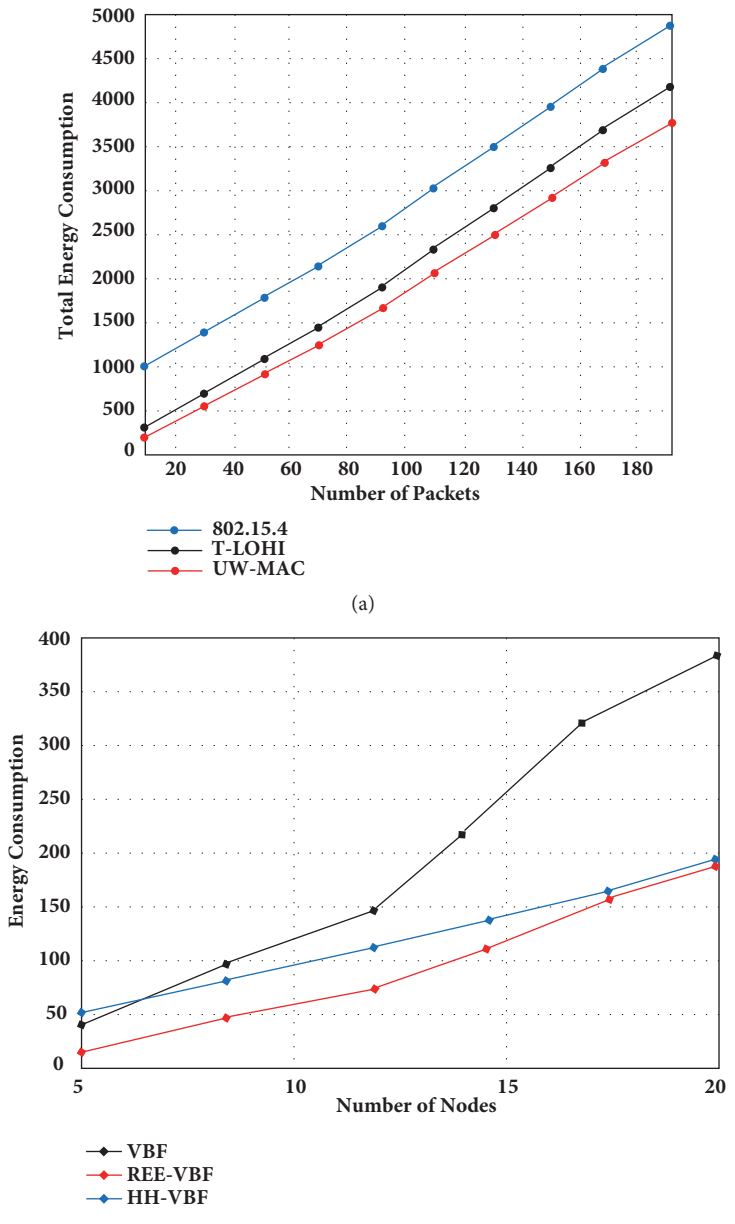


Figure 6: (a) Energy consumption of T-Lohi and MACUWASN [17]. (b) Energy consumption of VBF, REE-VBF, and HH-VBF [17].

routing to make it more energy efficient than VBF. On the other hand, instead of flooding, optimal relay is selected in REE-VBF due to which the performances of REE-VBF and HH-VBF have passed the similar points.

In Figure 7, different constraints are shown which are considered while designing the routing protocols for underwater sensor networks (UWSNs); these constraints may vary according to situation and requirement for a scenario.

6. Conclusion and Future Directions

In this paper, we have discussed several techniques of underwater sensor networks. The objective of the reviewed techniques is to overcome the underwater challenges and to give directions to future researchers. Also, we presented a vibrant view to academia by providing a base for a better solution. In this perspective, we have presented future directions which are still not yet explored in this research area. A better communication technique can be proposed by considering environmental effect during communication. In the development of underwater communication technique utmost care must be taken regarding the life of marine animals and their communication. The deep digging out in the areas regarding nonlinear sound propagation of acoustic signals can be more useful for designing future communication techniques. The future identified research areas include cognitive networks area and underwater spectrum for their efficient use. Major challenges for the design of cognitive acoustic network

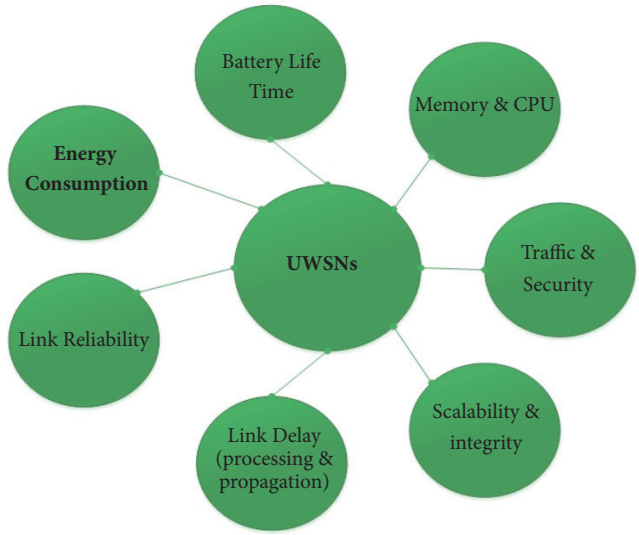


Figure 7: UWSNs constraints for designing routing protocols.

Table 7: Comparison MCA protocols.

Protocol	Based	Topology	Advantages	Disadvantages
T-Lohi	Contention	Not fixed/Distributed	It is used to solve the space time & uncertain data reservation problem while using short wake up time to reduce energy consumption.	Node is required to be idle and listen to the channel each in contention round.
R-MAC	Contention	Multi-hop	TC & data packets are scheduled to the both sender and receiver side.	There is No technique to join for new nodes when node dies or left.
H-MAC	Hybrid	Centralized	It accepts the benefits from both contentions based &contention-free protocols while low power consumption/energy reservation.	Not Optimal for dense & heavily loaded network.
UWAN-MAC	Hybrid	Dense network	It achieves synchronized locally schedule even when long propagation delay is present because it required small duty cycle.	Difficult to achieve high throughput due to small duty cycle feature.
P-MAC	Hybrid	Centralized	According to information of VDL it works dynamically & adaptively.	The addition of P-MAC with multichannel& ad-hoc mechanisms.
MFAMA	Contention	Mobile UWASN	A greedy approach intends to maximize throughput.	pay compensation to the Fairness.

are spectrum sensing, dynamic power control, spectrum sensing strategy, spectrum sharing, and spectrum decision. Therefore, routing and media access control protocols need to design by taking care of maximizing channel utilization. Standardization is highly required at media access control level. In addition, standards for underwater sensor networks are another challenge for academia. Further investigation is required in localization for underwater sensor networks. A GPS like localization sachem is still not created for underwater sensor networks and localization of a freely moving node is still an open area for research. Besides this, variable length packet in communication can further be investigated. Thus, we intend to target in future the variable packet size selection to improve the utilization of acoustic channel.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Soonchunhyang University Research Fund and the MSIP (Ministry of Science, ICT and

Future Planning), Korea, grant number IITP-2018-2014-1-00720 and the APC was funded by IITP-2018-2014-1-00720.

References

[1] N.-S. N. Ismail, L. A. Hussein, and S. H. S. Ariffin, "Analyzing the performance of acoustic channel in underwater wireless sensor network(UWSN)," in *Proceedings of the Asia Modelling Symposium 2010: 4th International Conference on Mathematical Modelling and Computer Simulation, AMS2010*, pp. 550–555, Malaysia, May 2010.

[2] T. Cinar and M. Bulent Orencik, "An underwater acoustic channel model using ray tracing in ns-2," in *Proceedings of the 2009 2nd IFIP Wireless Days (WD 2009)*, pp. 1–6, Paris, December 2009.

[3] J. Lloret, "Underwater sensor nodes and networks," *Sensors*, vol. 13, no. 9, pp. 11782–11796, 2013.

[4] M. Ayaz and A. Abdullah, "Underwater wireless sensor networks: Routing issues and future challenges," in *Proceedings of 7th International Conference on Advances in Mobile Computing and Multimedia, MoMM2009*, pp. 370–375, Malaysia, December 2009.

[5] L. Liu, S. Zhou, and J. H. Cui, "Prospects and problems of wireless communication for underwater sensor networks," *Wireless Communications and Mobile Computing*, vol. 8, no. 8, pp. 977–994, 2008.

[6] G. Zaibi, N. Nasri, A. Kacouri, and M. Samet, "Survey of temperature variation effect on underwater acoustic wireless transmission," *ICGST-CNIR Journal*, vol. 9, 2009.

[7] Y. Luo, L. Pu, M. Zuba, Z. Peng, and J.-H. Cui, "Challenges and opportunities of underwater cognitive acoustic networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 2, pp. 198–211, 2014.

[8] X.-P. Gu, Y. Yang, and R.-L. Hu, "Analyzing the performance of channel in Underwater Wireless Sensor Networks(UWSN)," in *Proceedings of the 2011 International Conference on Advanced in Control Engineering and Information Science, CEIS 2011*, pp. 95–99, China, August 2011.

[9] J. Kong, M. Gerla, and S. Zhou, *Challenges: Building Scalable and Distributed Underwater Wireless Sensor Networks (UWSNs) for Aquatic Applications*, 2005.

[10] S. Basagni, C. Petrioli, R. Petroccia, and M. Stojanovic, "Optimized packet size selection in underwater wireless sensor network communications," *IEEE Journal of Oceanic Engineering*, vol. 37, no. 3, pp. 321–337, 2012.

[11] S. Basagni, C. Petrioli, R. Petroccia, and M. Stojanovic, "Optimizing network performance through packet fragmentation in multi-hop underwater communications," in *Proceedings of the OCEANS'10 IEEE Sydney, OCEANSSYD 2010*, pp. 1–7, Australia, May 2010.

[12] S. Basagni, C. Petrioli, R. Petroccia, and M. Stojanovic, "Choosing the packet size in multi-hop underwater networks," in *Proceedings of the OCEANS'10 IEEE Sydney, OCEANSSYD 2010*, pp. 1–9, Australia, May 2010.

[13] M. Ayaz, L. T. Jung, A. Abdullah, and I. Ahmad, "Reliable data deliveries using packet optimization in multi-hop underwater sensor networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 24, no. 1, pp. 41–48, 2012.

[14] N. Li, J.-F. Martínez, J. M. M. Chaus, and M. Eckert, "A survey on underwater acoustic sensor network routing protocols," *Sensors*, vol. 16, no. 3, p. 414, 2016.

[15] M. Faheem, G. Tuna, and V. C. Gungor, "QERP: Quality-of-Service (QoS) Aware Evolutionary Routing Protocol for Underwater Wireless Sensor Networks," *IEEE Systems Journal*, 2017.

[16] C. Zidi, F. Bouabdallah, and R. Boutaba, "Routing design avoiding energy holes in underwater acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 16, no. 14, pp. 2035–2051, 2016.

[17] N. Z. Zenia, M. Aseeri, M. R. Ahmed, Z. I. Chowdhury, and M. Shamim Kaiser, "Energy-efficiency and reliability in MAC and routing protocols for underwater wireless sensor network: A survey," *Journal of Network and Computer Applications*, vol. 71, pp. 72–85, 2016.

[18] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[19] I. F. Akyildiz, D. Pompili, and T. Melodia, "Challenges for efficient communication in underwater acoustic sensor networks," *SIGBED Review*, vol. 1, no. 2, pp. 3–8, 2004.

[20] C. Peach and A. Yarali, "An Overview of Underwater Sensor Networks," in *Proceedings of the routing techniques regarding network layer*, pp. 31–36, 2013.

[21] I. F. Akyildiz, D. Pompili, and T. Melodia, "State-of-the-art in protocol research for underwater acoustic sensor networks," in *Proceedings of the 1st ACM International Workshop on Underwater Networks*, pp. 7–16, ACM, September 2006.

[22] D. H. Cato, "Ocean ambient noise: Its measurement and its significance to marine animals," in *Proceedings of the Conference on Underwater Noise Measurement, Impact and Mitigation*, pp. 1–9, Institute of Acoustics, Southampton, UK, 2008.

[23] E. J. Harland, SAS. Jones, and T. Clarke, "SEA 6 Technical report: Underwater ambient noise," A report by QinetiQ as part of the UK Department of Trade and Industrys offshore energy Strategic Environmental Assessment programme, 2005.

[24] M. Stojanovic, "Underwater acoustic communications: design considerations on the physical layer," in *Proceedings of the 5th Annual Conference on Wireless on Demand Network Systems and Services (WONS '08)*, pp. 1–10, January 2008.

[25] K. A. Perrine, K. F. Nieman, T. L. Henderson, K. H. Lent, T. J. Brudner, and B. L. Evans, "Doppler estimation and correction for shallow underwater acoustic communications," in *Proceedings of the 44th Asilomar Conference on Signals, Systems and Computers, Asilomar 2010*, pp. 746–750, USA, November 2010.

[26] X. Zhang, X. Han, J. Yin, and X. Sheng, "Study on Doppler effects estimate in underwater acoustic communication," in *Proceedings of the ICA 2013 Montreal*, pp. 070062-070062, Montreal, Canada, 2013.

[27] M. Garcia, S. Sendra, G. Lloret, and J. Lloret, "Monitoring and control sensor system for fish feeding in marine fish farms," *IET Communications*, vol. 5, no. 12, pp. 1682–1690, 2011.

[28] C. Lv, S. Wang, M. Tan, and L. Chen, "UA-MAC: An underwater acoustic channel access method for dense mobile underwater sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 374028, 2014.

[29] S. Shahabudeen, M. Chitre, and M. Motani, "Adaptive multimode medium access control for underwater acoustic networks," *IEEE Journal of Oceanic Engineering*, vol. 39, no. 3, pp. 500–514, 2014.

[30] A. Sehgal, I. Tumar, and J. Schönwälder, "Variability of available capacity due to the effects of depth and temperature in the underwater acoustic communication channel," in *Proceedings*

of the OCEANS '09 IEEE Bremen: Balancing Technology with Future Needs, Germany, May 2009.

[31] A. F. Harris III, D. G. B. Meneghetti, and M. Zorzi, "Maximizing channel utilization for underwater acoustic links," in *Proceedings of the OCEANS 2007 - Europe*, pp. 1–6, June 2007.

[32] Y. Guo and Y. Liu, "Localization for anchor-free underwater sensor networks," *Computers and Electrical Engineering*, vol. 39, no. 6, pp. 1812–1821, 2013.

[33] X.-J. Du, K.-J. Huang, S.-L. Lan, Z.-X. Feng, and F. Liu, "LB-AGR: level-based adaptive geo-routing for underwater sensor network," *The Journal of China Universities of Posts and Telecommunications*, vol. 21, no. 1, pp. 54–59, 2014.

[34] C.-J. Huang, Y.-W. Wang, H.-H. Liao, C.-F. Lin, K.-W. Hu, and T.-Y. Chang, "A power-efficient routing protocol for underwater wireless sensor networks," *Applied Soft Computing*, vol. 11, no. 2, pp. 2348–2355, 2011.

[35] T. Ali, L. T. Jung, and I. Faye, "Diagonal and Vertical Routing Protocol for Underwater Wireless Sensor Network," *Procedia - Social and Behavioral Sciences*, vol. 129, pp. 372–379, 2014.

[36] J. Lloret, S. Sendra, M. Ardid, and J. J. P. C. Rodrigues, "Underwater wireless sensor communications in the 2.4 GHz ISM frequency band," *Sensors*, vol. 12, no. 4, pp. 4237–4264, 2012.

[37] B. Z. Chen and D. Pompili, "Reliable geocasting for random-access underwater acoustic sensor networks," *Ad Hoc Networks*, vol. 21, pp. 134–146, 2014.

[38] L. T. Jung and A. B. Abdullah, "Underwater wireless network: Energy efficiency and optimal data packet size," in *Proceedings of the 1st International Conference on Electrical, Control and Computer Engineering 2011, InECCE 2011*, pp. 178–182, Malaysia, June 2011.

[39] S. Gopi, K. Govindan, D. Chander, U. B. Desai, and S. N. Merchant, "E-PULRP: energy optimized path unaware layered routing protocol for underwater sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3391–3401, 2010.

[40] T. Hu and Y. Fei, "QELAR: a machine-learning-based adaptive routing protocol for energyefficient and lifetime-extende d underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 6, pp. 796–809, 2010.

[41] M. Ayaz, A. Abdullah, I. Faye, and Y. Batira, "An efficient dynamic addressing based routing p rotocol for underwater wireless sensor networks," *Computer Communications*, vol. 35, no. 4, pp. 475–486, 2012.

[42] H. Yan, Z. J. Shi, and J.-H. Cui, "DBR: Depth-Based Routing for underwater sensor networks," in *NETWORKING 2008 Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet*, vol. 4982 of *Lecture Notes in Computer Science*, pp. 72–86, Springer, 2008.

[43] M. R. Jafri, M. M. Sandhu, K. Latif, Z. A. Khan, A. U. H. Yasar, and N. Javaid, "Towards delay-sensitive routing in underwater wireless sensor networks," in *Proceedings of 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks*, vol. 37, pp. 228–235, Canada, September 2014.

[44] S. Zhang, D. Li, and J. Chen, "A link-state based adaptive feedback routing for underwater acoustic sensor networks," *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4402–4412, 2013.

[45] Z. Guo, G. Colombi, B. Wang, J. H. Cui, D. Maggiorinit, and G. P. Rossi, "Adaptive routing in underwater delay/disruption tolerant sensor networks," in *Proceedings of the 5th Annual Conference on Wireless on Demand Network Systems and Services (WONS '08)*, pp. 31–39, January 2008.

[46] Y.-S. Chen and Y.-W. Lin, "Mobicast routing protocol for underwater sensor networks," *IEEE Sensors Journal*, vol. 13, no. 2, pp. 737–749, 2013.

[47] Z. Li, N. Yao, and Q. Gao, "Relative Distance Based Forwarding Protocol for Underwater Wireless Networks," *International Journal of Distributed Sensor Networks*, pp. 1–11, 2014.

[48] S. K. Dhurandher, M. S. Obaidat, and M. Gupta, "A novel geocast technique with hole detection in underwater sensor networks," in *Proceedings of 2010 ACS/IEEE International Conference on Computer Systems and Applications, AICCSA 2010*, Tunisia, May 2010.

[49] Y. Noh, U. Lee, P. Wang, B. S. C. Choi, and M. Gerla, "VAPR: void-aware pressure routing for underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, pp. 895–908, 2013.

[50] S. Basagni, C. Petrioli, R. Petroccia, and D. Spaccini, "Channel-aware routing for underwater wireless networks," in *Proceedings of the OCEANS 2012 MTS/IEEE - YEOSU*, pp. 1–9, May 2012.

[51] A. Wahid and D. Kim, "An Energy Efficient Localization-Free Routing Protocol for Underwater Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 307246, 11 pages, 2012.

[52] F. Thomsen, "Assessment of the environmental impact of underwater noise," *OSPAR Commission. Biodiversity Series*, 2009.

[53] A. Rahman, V. Muthukkumarasamy, and E. Sithirasenan, "The analysis of temperature, depth, salinity effect on acoustic speed for a vertical water column," in *Proceedings of the 9th IEEE International Conference on Distributed Computing in Sensor Systems, DCoSS 2013*, pp. 310–312, USA, May 2013.

[54] M. Garcia-Pineda, S. Sendra, M. Atenas, and J. Lloret, "Underwater wireless ad-hoc networks: A survey," *Mobile ad hoc Networks: Current Status and Future Trends*, pp. 379–411, 2011.

[55] P. P. Kumar, R. Bhagat, and S. Suvarna, *Underwater Channel Simulation*, PRATHEEK, 2017.

[56] K. V. Mackenzie, "Nine-term equation for sound speed in the oceans," *The Journal of the Acoustical Society of America*, vol. 70, no. 3, pp. 807–812, 1981.

[57] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007.

[58] H. Jindal, S. Saxena, and S. Singh, "Challenges and issues in underwater acoustics sensor networks: A review," in *Proceedings of the 2014 3rd IEEE International Conference on Parallel, Distributed and Grid Computing, PDGC 2014*, pp. 251–255, India, December 2014.

[59] E. Felemban, F. K. Shaikh, U. M. Qureshi, A. A. Sheikh, and S. B. Qaisar, "Underwater sensor network applications: a comprehensive survey," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 896832, 2015.

[60] H.-P. Tan, R. Diamant, W. K. G. Seah, and M. Waldmeyer, "A survey of techniques and challenges in underwater localization," *Ocean Engineering*, vol. 38, no. 14-15, pp. 1663–1676, 2011.

[61] M. Khalid, Z. Ullah, N. Ahmad et al., "A survey of routing issues and associated protocols in underwater wireless sensor networks," *Journal of Sensors*, vol. 2017, Article ID 7539751, 17 pages, 2017.

[62] M. Ayaz, I. Baig, A. Abdullah, and I. Faye, "A survey on routing techniques in underwater wireless sensor networks," *Journal of Network and Computer Applications*, vol. 34, no. 6, pp. 1908–1927, 2011.

[63] G. Han, J. Jiang, N. Bao, L. Wan, and M. Guizani, "Routing protocols for underwater wireless sensor networks," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 72–78, 2015.

[64] E. M. Sozer, M. Stojanovic, and J. G. Proakis, "Underwater acoustic networks," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1, pp. 72–82, 2000.

[65] M. P. Malumbres, P. P Garrido, C. T. Calafate, and J. Oliver, *Underwater Wireless Networking Technologies*, Miguel Hernandez University, Spain Technical University of Valencia, Valencia, Spain, 2009.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 9710656, 7 pages
https://doi.org/10.1155/2019/9710656

Research Article

An Angle Estimation Method for Monostatic MIMO Radar Based on RCC-FLOM Algorithm

Jian Gong,[1,2,3] **Huan Wang**,[2] **and Yiduo Guo**[2]

[1] *University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China*
[2] *Air Force Engineering University, Xi'an, Shaanxi 710051, China*
[3] *TongFang Electronic Science and Technology Co. Ltd., Jiujiang Jiangxi 332007, China*

Correspondence should be addressed to Jian Gong; wcyls@163.com

Received 20 September 2018; Revised 25 November 2018; Accepted 11 December 2018; Published 1 January 2019

Guest Editor: Fawad Zaman

The performance of the angle estimation algorithm based on the two-order or higher order cumulants in the impact noise background will decline sharply. Therefore, it is necessary to study the new algorithm to estimate target angle in the impact noise background. In order to solve the angle estimation problem of coherent sources in the impulse noise background, a conjugate rotation invariant subspace algorithm based on reduced order fractional lower order covariance matrix is proposed. Use the reduced dimension lower order fraction covariance matrix to reduce the impulse noise influence. And according to the conjugate rotation invariant subspace, the coherent source is decohered. The Monte-Carlo experiments show that the proposed algorithm has the advantages of high estimation probability and low root mean square error in the case of low signal-to-noise ratio, compared with the existing FLOM-MUSIC algorithm and FLOM-Unitary ESPRIT algorithm.

1. Introduction

The noise encountered in the practical application of Multiple-Input Multiple-Output (MIMO) radar often has certain impact characteristics, such as cosmic noise, atmospheric noise, and meteorological noise. In recent years, a large number of experimental data and simulation results [1–3] confirm that the impulse noise is consistent with the α stable distribution. Because the α stable distribution does not have two-order square or higher, this means that the performance of the angle estimation algorithm based on the two-order or higher order cumulants in the impact noise background will decline sharply. Therefore, it is necessary to study the new algorithm to estimate target angle in the impact noise background.

Tsakalides [4–6] and Liu [7] estimate the direction of arrival using Multiple Signal Classification (MUSIC) algorithm based on covariance and fractional lower order moment respectively, but all of them need to search spectral peak, and the amount of operation and storage is very large. Lv Zejun [8] and He Jin [9] estimate the arrival angle based on covariance in the impact noise environment, but the above methods need to estimate the appropriate fractional lower order parameters. Li Li [10] uses the Maximum Correntropy Criterion (MCC) to derive the ductile parallel factor algorithm for the impact noise environment, but has not solved the coherent source problem.

Based on the above research, we combine reduced dimension conjugate covariance (RCC) with fraction lower order matrix (FLOM) and then propose a RCC-FLOM algorithm for monostatic MIMO radar. The proposed algorithm can reduce the impulse noise influence and has the advantages of high estimation probability. The remainder of this paper is organized as follows. In Section 2, the monostatic MIMO radar echo model is established. Then, the angle estimation method based on RCC-FLOM is proposed in Section 3. Some simulations are conducted to verify the performance of the proposed method in Section 4. Finally, we conclude the paper in Section 5.

Notation: $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and conjugate-transpose operators; $\otimes$ denotes the Kronecker product; $E[\cdot]$ denotes the expected value; $(\cdot)^*$ denotes the conjugate operator; $\sum(\cdot)$ denotes the sum operator.

2. Monostatic MIMO Radar Echo Model

Considering the monostatic MIMO radar as shown in Figure 1, the transmitting and receiving antenna adopt M and N elements, respectively. The spacing of array elements is $d = \lambda/2$, λ is carrier wavelength. Assuming that the direction of arrival (DOA) and the direction of departure (DOD) both are θ_p. The echo signal processed by matching filtering as below,

$$\mathbf{y}(t) = \mathbf{A}\boldsymbol{\beta}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_r(\theta_1)\otimes\mathbf{a}_t(\theta_1), \mathbf{a}_r(\theta_2)\otimes\mathbf{a}_t(\theta_2), \cdots, \mathbf{a}_r(\theta_P)\otimes\mathbf{a}_t(\theta_P)]$, the transmit steering vector of the *pth* target is $\mathbf{a}_t(\theta_p) = [1, \exp(-j\pi\sin\theta_p), \cdots, \exp(-j\pi(M-1)\sin\theta_p)]^{\mathrm{T}}$, and the receive steering vector is $\mathbf{a}_r(\theta_p) = [1, \exp(-j\pi\sin\theta_p), \cdots, \exp(-j\pi(N-1)\sin\theta_p)]^{\mathrm{T}}$, $\boldsymbol{\beta} = [\xi_1 e^{j2\pi f_{d1}t}, \xi_2 e^{j2\pi f_{d2}t}, \cdots, \xi_P e^{j2\pi f_{dP}t}]^{\mathrm{T}}$, ξ_p is the reflection coefficient of the *pth* target, f_{dp} is the normalized Doppler frequency of the *pth* target, and $\mathbf{n}(t)$ representation of noisy column vectors.

3. Angle Estimation Based on RCC-FLOM

3.1. Reduced Dimension Lower Order Fraction Covariance Matrix. If defined,

$$\mathbf{a}_r(\theta_p)\otimes\mathbf{a}_t(\theta_p) \triangleq \mathbf{F}\mathbf{b}(\theta_p) \tag{2}$$

In the form, $\mathbf{b}(\theta_p) = [1, \exp(-j\pi\sin\theta_p), \cdots, \exp(-j\pi(M+N-1)\sin\theta_p)]^{\mathrm{T}}$ and $\mathbf{F} \in \mathbf{C}^{MN\times(M+N-1)}$ is a dimensionality reduction matrix, which can be expressed as

$$\mathbf{F} = \begin{bmatrix} \left.\begin{matrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{matrix}\right\} M \\ \left.\begin{matrix} 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \end{matrix}\right\} M \\ \vdots \\ \left.\begin{matrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{matrix}\right\} M \end{bmatrix} \in \mathbf{C}^{MN\times(M+N-1)} \tag{3}$$

Then the matrix $\mathbf{A}$ can be expressed as

$$\mathbf{A} = \mathbf{F}\mathbf{A}' \tag{4}$$

where $\mathbf{A}' = [\mathbf{a}'(\theta_1), \mathbf{a}'(\theta_2), \cdots, \mathbf{a}'(\theta_P)] \in \mathbf{C}^{(M+N-1)P}$, $\mathbf{a}'(\theta_p) = [1, a_p, \cdots, a_p^{I-1}]^{\mathbf{T}}$, $a_p = e^{j\pi\sin\theta_p}$, and it is a Vanermonde matrix. I is the virtual elements number after dimension reduction transformation, then $I = M + N - 1$.

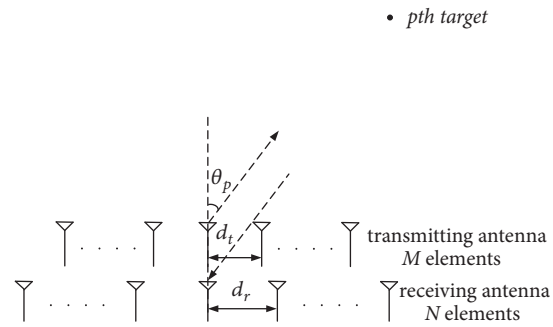


Figure 1: The structure of a monostatic MIMO radar.

According to formula (3), if formula $\mathbf{W} \triangleq \mathbf{F}^{\mathrm{H}}\mathbf{F}$ is defined, there are

$$\mathbf{W} = \mathrm{diag}\left(1, 2, \cdots, \underbrace{\min(M,N), \cdots, \min(M,N)}_{|M-N|+1}, \cdots, 2, 1\right) \tag{5}$$

If both sides of formula (1) are multiplied by $\mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$, we can get

$$\mathbf{y}' = \mathbf{A}'\boldsymbol{\beta} + \mathbf{n}' \tag{6}$$

where $\mathbf{n}'(t_l) = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}\mathbf{n}(t_l)$ is a $(M+N-1)$ dimension additive noise vector.

When $\mathbf{n}'$ is impact noise, $\mathbf{y}'$ has no more than two-order square. According to [6, 7], the FLOM of $\mathbf{y}'$ can be constructed as

$$C_{uv} = E\left[y_u'(t)\left|y_v'(t)\right|^{p-2} y_v'^{*}(t)\right] \quad 1 < p < \alpha \le 2 \tag{7}$$

where $y_u'(t)$ and $y_v'(t)$ is the output of matched filtering after MIMO radar dimensionality reduction. C_{uv} is the (u, v) element of the fractional lower order covariance matrix $\mathbf{C}$ and satisfies $1 \le u, v \le I$.

Substituting formula (6) in formula (7), C_{uv} can be expressed as

$$\begin{aligned} C_{uv} = {} & \sum_{v=1}^{P} A'_{uv} \\ & \cdot E\left[\beta_v\left|\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v\right|^{p-2}\left(\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v\right)^{*}\right] \\ & + E\left[n_u\left|\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v\right|^{p-2}\left(\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v\right)^{*}\right] \end{aligned} \tag{8}$$

It can be written as a matrix as follows:

$$\mathbf{C} = \mathbf{A}'\boldsymbol{\Lambda}\mathbf{A}'^{\mathrm{H}} + \gamma\mathbf{I} \tag{9}$$

where $\Lambda_{uv} = \delta_{uv}E[\beta_v|\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v|^{p-2}(\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v)^{*}]$ and $\gamma = E[n_u|\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v|^{p-2}(\sum_{q=1}^{P} A'_{uv}\beta_q + n'_v)^{*}]$.

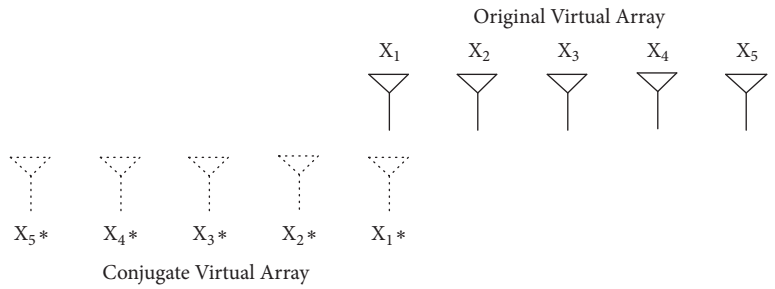


Figure 2: Monostatic MIMO radar conjugate virtual array.

After obtaining the reduced order fractional lower order covariance matrix formula (9), the eigenvalues of $\mathbf{C}$ according to [11, 12] are as follows:

$$\mathbf{C} = \mathbf{E}_S \mathbf{\Lambda}_S \mathbf{E}_S{}^{\mathrm{H}} + \mathbf{E}_N \mathbf{\Lambda}_N \mathbf{E}_N{}^{\mathrm{H}} \tag{10}$$

where $\mathbf{E}_S$ is a subspace formed by eigenvectors corresponding to large eigenvalues, i.e., signal subspace, while $\mathbf{E}_N$ is a subspace formed by eigenvectors corresponding to small eigenvalues, i.e., noise subspace.

Under ideal conditions, the signal subspace and the noise subspace are orthogonal to each other; that is, the steering vector in the signal subspace is also orthogonal to the noise subspace as

$$\mathbf{a}'^{\mathrm{H}}(\theta)\,\mathbf{E}_N = 0 \tag{11}$$

The classical MUSIC algorithm is based on the above properties. However, considering the fact that the actual received data matrix is finite and there is noise, $\mathbf{a}'(\theta)$ and $\mathbf{E}_N$ can not be completely orthogonal; that is to say, formula (11) does not hold. Therefore, in fact, DOA is realized by minimum optimization search; that is,

$$\theta_{MUSIC} = \arg\min_{\theta} \mathbf{a}'^{\mathrm{H}}(\theta)\,\mathbf{E}_N \mathbf{E}_N^{\mathrm{H}} \mathbf{a}'(\theta) \tag{12}$$

So the target angle can be estimated by the MUSIC algorithm's spectral search formula such as

$$P_{\mathrm{MUSIC}}(\theta) = \frac{1}{\mathbf{a}'^{\mathrm{H}}(\theta)\,\mathbf{E}_N \mathbf{E}_N^{\mathrm{H}} \mathbf{a}'(\theta)} \tag{13}$$

3.2. Conjugate Rotation Invariant Subspace Algorithm. Constructing the lower order fractional lower order covariance matrix can only solve the problem of impulse noise. When the multipath effect and electronic interference are affected, the angle estimation of coherent sources is also needed. In this paper, the conjugate rotation invariant subspace algorithm (C-ESPRIT) is used to solve the problem of coherent source angle estimation. Taking the MIMO radar with 5 virtual elements as an example, the virtual space array constructed is shown in Figure 2.

The target echo data obtained by virtual subarray 1 can be expressed as follows:

$$\mathbf{y}_{e1} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_P \\ y_2 & y_3 & y_4 & \cdots & y_{P+1} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{I-P} & y_{I-P+1} & y_{I-P+2} & \cdots & y_{I-1} \\ y_I^* & y_{I-1}^* & y_{I-2}^* & \cdots & y_{I-P+1}^* \\ \vdots & \vdots & \vdots & & \vdots \\ y_{P+2}^* & y_{P+1}^* & y_P^* & \cdots & y_3^* \\ y_{P+1}^* & y_P^* & y_{P-1}^* & \cdots & y_2^* \end{bmatrix}_{2(I-P)\times P} \tag{14}$$

According to formula (6), the matrix $\mathbf{y}_{e1}$ can be expressed as

$$\mathbf{y}_{e1} = \mathbf{R}\mathbf{A}'(\theta) + \mathbf{n}_{e1} \tag{15}$$

where

$$\mathbf{R} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_P \\ \beta_1 a_1 & \beta_2 a_2 & \beta_3 a_3 & \cdots & \beta_P a_P \\ \vdots & \vdots & \vdots & & \vdots \\ \beta_1 a_1^{I-P-1} & \beta_2 a_2^{I-P-1} & \beta_3 a_3^{I-P-1} & \cdots & \beta_P a_P^{I-P-1} \\ \beta_1^* a_1^{-(I-1)} & \beta_2^* a_2^{-(I-1)} & \beta_3^* a_3^{-(I-1)} & \cdots & \beta_P^* a_P^{-(I-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ \beta_1^* a_1^{-(P+1)} & \beta_2^* a_2^{-(P+1)} & \beta_3^* a_3^{-(P+1)} & \cdots & \beta_P^* a_P^{-(P+1)} \\ \beta_1^* a_1^{-P} & \beta_2^* a_2^{-P} & \beta_3^* a_3^{-P} & \cdots & \beta_P^* a_P^{-P} \end{bmatrix}_{2(I-P)\times P} \tag{16}$$

The target echo data obtained by virtual subarray 2 can be expressed as follows:

$$\mathbf{y}_{e2} = \begin{bmatrix} y_2 & y_3 & y_4 & \cdots & y_{P+1} \\ y_3 & y_4 & y_5 & \cdots & y_{P+2} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{I-P+1} & y_{I-P+2} & y_{I-P+3} & \cdots & y_I \\ y_{I-1}^* & y_{I-2}^* & y_{I-3}^* & \cdots & y_{I-P}^* \\ \vdots & \vdots & \vdots & & \vdots \\ y_{P+1}^* & y_P^* & y_{P-1}^* & \cdots & y_2^* \\ y_P^* & y_{P-1}^* & y_{P-2}^* & \cdots & y_1^* \end{bmatrix}_{2(I-P)\times P} \tag{17}$$

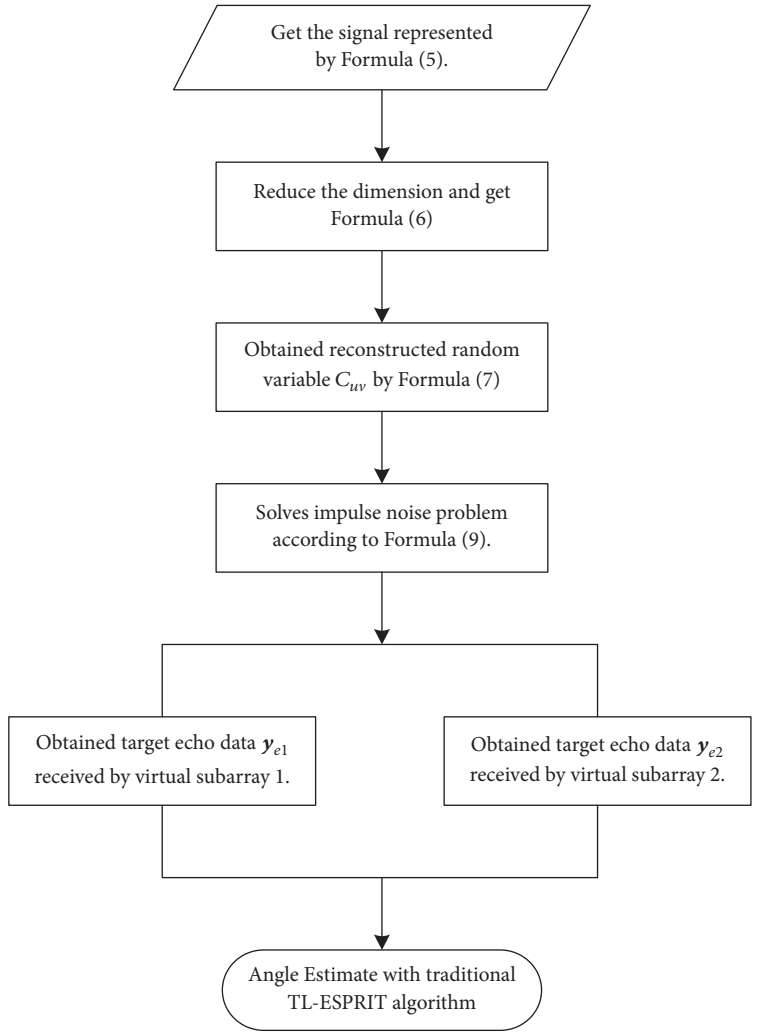


Figure 3: Flow chart of the RCC-FLOM algorithm.

According to formula (6), the matrix $\mathbf{y}_{e2}$ can be expressed as

$$\mathbf{y}_{e2} = \mathbf{R}\boldsymbol{\Phi}\mathbf{A}'(\theta) + \mathbf{n}_{e2} \tag{18}$$

where $\boldsymbol{\Phi}$ is a rotation factor and its expression is $\boldsymbol{\Phi} = \text{diag}(a_1, a_2, \cdots, a_P)$.

The matrix $\mathbf{A}'(\theta)$ is a Vanermonde matrix, because the coherent source is located in different directions of space and linearly independent of each row, $\text{rank}(\mathbf{A}'(\theta)) = P$.

When $2(I-P) \geq P$, the $\mathbf{R}$ of the provable matrix is linearly independent; that is, $\text{rank}(\mathbf{R}) = P$. Because of $\text{rank}(\boldsymbol{\Phi}) = P$, the rank of virtual subarray $\mathbf{y}_{e1}$ and $\mathbf{y}_{e2}$ is P. Therefore, the traditional TL-ESPRIT algorithm can be used to estimate P coherent sources.

3.3. The Steps of the RCC-FLOM Algorithm. Based on the above analysis, the operation steps of the RCC-FLOM algorithm can be summarized as follows.

Step 1. The echo signal of single pulse is matched filtered to get the signal represented by formula (5).

Step 2. Formula (6) is obtained for reducing the dimension of the signal represented by formula (5).

Step 3. According to formula (7), the fractional lower order covariance $\mathbf{y}'$ of the reconstructed random variable C_{uv} is obtained.

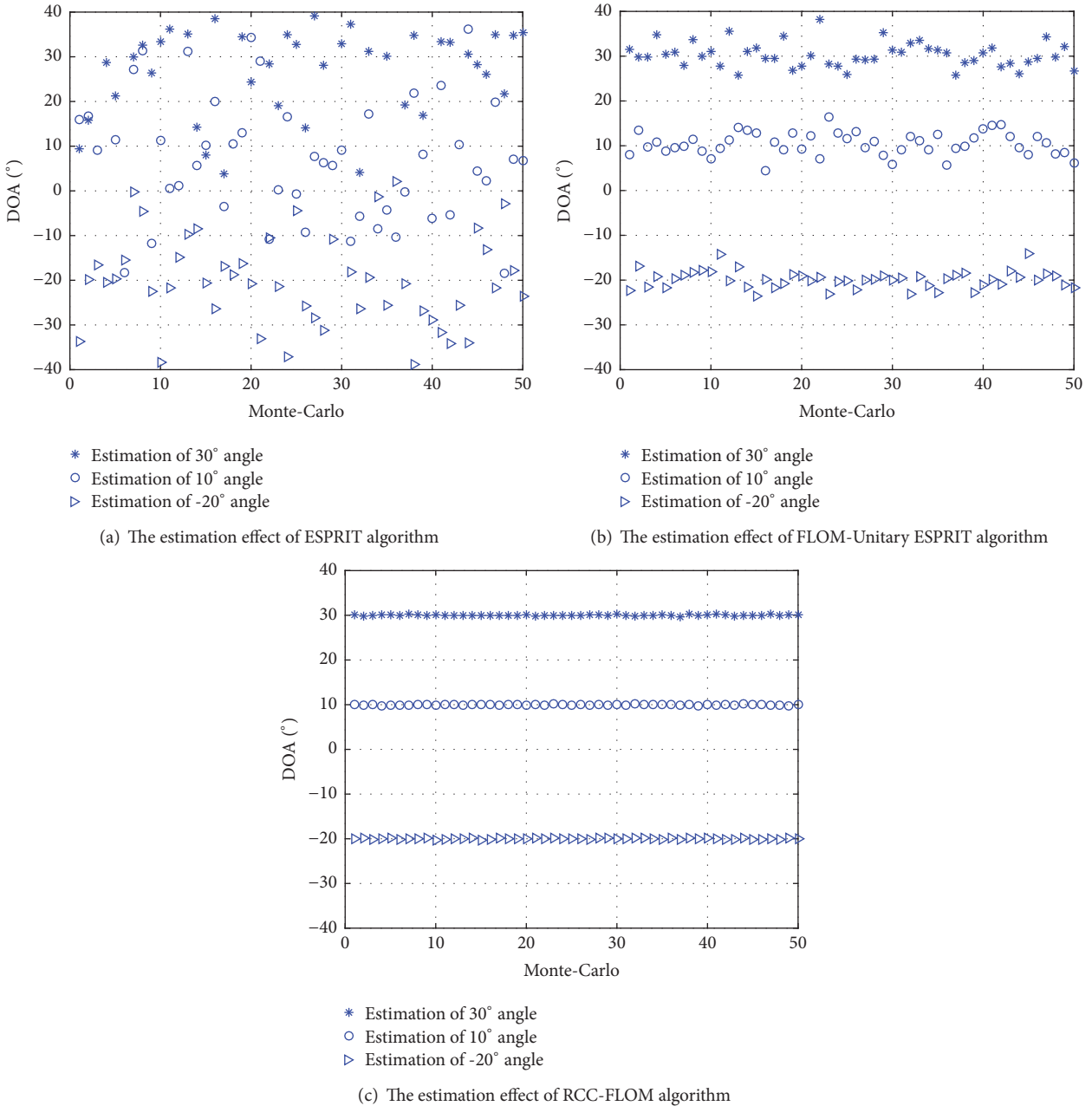
Step 4. The fractional lower order covariance C_{uv} is rewritten into the matrix form shown in formula (9), which solves the problem of impulse noise.

Step 5. According to formula (15), the target echo data $\mathbf{y}_{e1}$ received by virtual subarray 1 is obtained.

Step 6. According to formula (18), the target echo data $\mathbf{y}_{e2}$ received by virtual subarray 2 is obtained.

Step 7. Estimated the angle of P coherent sources based on the traditional TL-ESPRIT algorithm.

For a more intuitive presentation, the flow chart is drawn in Figure 3.



(a) The estimation effect of ESPRIT algorithm

(b) The estimation effect of FLOM-Unitary ESPRIT algorithm

(c) The estimation effect of RCC-FLOM algorithm

Figure 4: Angle estimation algorithms of coherent sources under impulse noise.

4. Computer Simulation Results

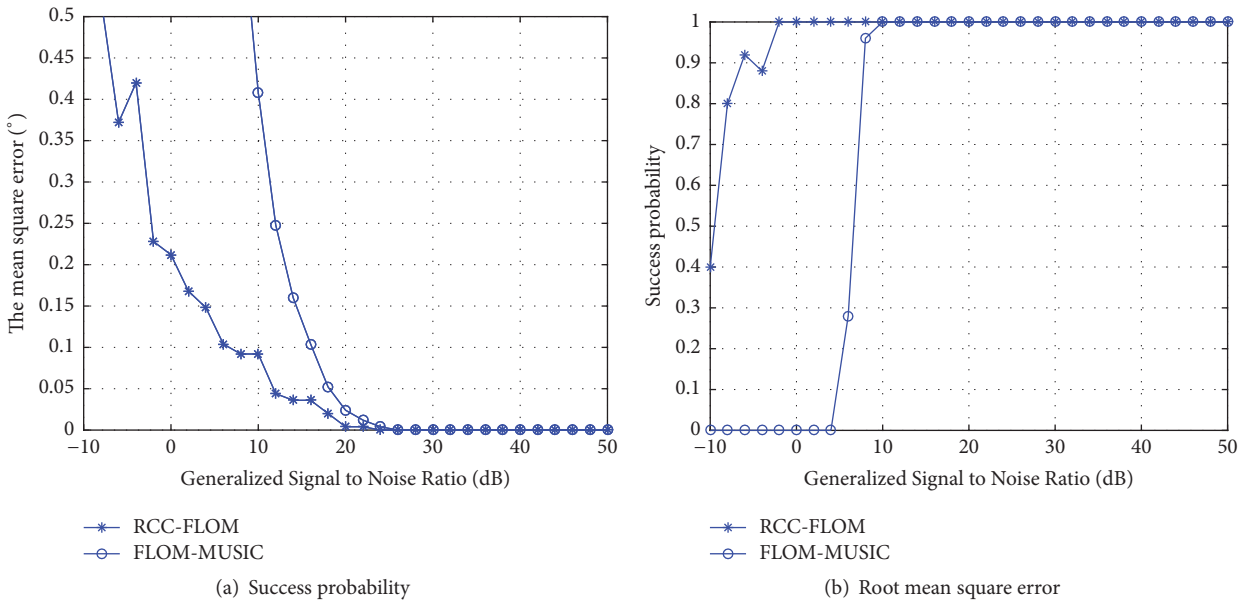
Assuming that the MIMO radar has 5 transmitting elements and 6 receiving elements, the number of virtual elements after dimension reduction is 10. The receiving and transmitting antennas are equal spaced uniform linear arrays with a spacing of 0.5λ. And we use MATLAB software to simulate the following experiments in this paper. The generalized signal-to-noise ratio (GSNR) is the ratio of average power to dispersion coefficient in simulation as follows:

$$GSNR = 10\lg\left(\frac{1}{\gamma N}\sum_{t=1}^{N}\left|s(t)^2\right|\right) \tag{19}$$

where N is the number of accumulated pulses. If $\alpha = 2$, the generalized signal-to-noise ratio is the same as the ordinary signal-to-noise ratio.

Experiment 1. Angle estimation of coherent sources under different impulse noise environments

Under the background of impact noise, the direction of arrival of 3 equal power coherent sources are 30°, 10°, −20°, respectively. The impulse noise of $S\alpha S$ distribution is $\alpha = 1.5$, and the generalized signal-to-noise ratio is 5 dB. Figures 4(a), 4(b), and 4(c), respectively, simulate the angle estimation of ESPRIT algorithm, FLOM-Unitary ESPRIT algorithm



(a) Success probability (b) Root mean square error

Figure 5: Performance comparison of different algorithms under impact noise.

proposed in the paper [13], and RCC-FLOM algorithm in 50 simulation experiments.

The simulation results of Figure 4 can be seen as follows: In the background of impulse noise, the coherent sources' direction of arrival can not be estimated accurately by using the ESPRIT algorithm. It is shown that the conventional subspace based algorithm can not suppress impulse noise and coherent decoherence when applied to monostatic MIMO radar. The RCC-FLOM algorithm proposed in this paper can accurately estimate the direction of arrival of coherent sources in the background of impulsive noise. The simulation results verify the effectiveness of the proposed algorithm.

Experiment 2. Comparison of statistical performance between the proposed RCC-FLOM algorithm and other algorithms

The mean square error (RSME) of DOA estimation is defined as

$$\mathrm{RMSE}(\theta) = \sqrt{\frac{1}{L_m}\sum_{l_m=1}^{L_m}\left(\widehat{\theta}_p - \theta_p\right)^2} \tag{20}$$

where L_m represents the number of Monte-Carlo experiments and θ_p and $\widehat{\theta}_p$ represent the true angle and estimated angle, respectively.

Under the background of impact noise, the direction of arrival of 2 equal power coherent sources is 30° and −20°, respectively. The impulse noise of $S\alpha S$ distribution is $\alpha = 1.5$. Figures 5(a) and 5(b), respectively, simulate the statistical performance of FLOM-MUSIC algorithm [11] and RCC-FLOM algorithm in 50 Monte-Carlo experiments.

The simulation results of Figure 5 can be seen as follows: Under the impact noise background, the RCC-FLOM algorithm proposed in this paper is superior to the FLOM-MUSIC algorithm proposed in the literature [11] in terms of the success probability and the root mean square error. With the increase of the generalized signal-to-noise ratio, the success probability of the two algorithms is gradually increased and the root mean square error is gradually reduced. That is, the performance of the angle estimation is increased with the increase of the generalized signal-to-noise ratio. When the SNR is high to a certain extent, the performance of the two algorithms is basically the same.

5. Conclusions

In this chapter, the MIMO radar angle estimation problem of coherent sources in the impulse noise background is studied. A conjugate rotation invariant subspace (RCC-FLOM) algorithm based on reduced order fractional lower order covariance matrix is proposed. The influence of impulse noise is reduced by using the method of constructing lower order fractional lower order covariance matrix. Then the coherent source is decohered according to the conjugate rotation invariant subspace method. It solves the problem that the two or four order statistical models can not effectively estimate the target angle under the impact noise background. The 50 Monte-Carlo experiments show that the proposed algorithm has the advantages of high estimation probability and low root mean square error in the case of low signal-to-noise ratio, compared with the existing FLOM-MUSIC algorithm.

Data Availability

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

Conflicts of Interest

No potential conflicts of interest were reported by the authors.

Authors' Contributions

Jian Gong and Yiduo Guo designed the algorithm scheme. Jian Gong performed the experiments and analysed the experiment results. Yiduo Guo and Huan Wang contributed to the manuscript drafting and critical revision. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants nos. 61601502 and 61501501.

References

[1] X. Ma and C. L. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2669–2687, 1996.

[2] P. Tsakalides, R. Raspanti, and C. L. Nikias, "Angle/doppler estimation in heavy-tailed clutter backgrounds," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 2, pp. 419–436, 1999.

[3] V. Katkovnik, "New concept of adaptive beamforming for moving sources and impulse noise environment," *Signal Processing*, vol. 80, no. 9, pp. 1863–1882, 2000.

[4] P. Tsakalides and C. L. Nikias, "Subspace-based direction finding in alpha-stable noise environments," in *Proceedings of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing*, pp. 2092–2095, IEEE, Shanghai, China, May 1995.

[5] P. Tsakalides and C. L. Nikias, "Maximum Likelihood Localization of Sources in Noise Modeled as a Stable Process," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2700–2713, 1995.

[6] P. Tsakalides and C. L. Nikias, "The robust covariation-based music (roc-music) algorithm for bearing estimation in impulsive noise environments," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1623–1633, 1996.

[7] T.-H. Liu and J. M. Mendel, "A subspace-based direction finding algorithm using fractional lower order statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1605–1613, 2001.

[8] L. Zejun and X. Xianci, "A novel method for estimation 2-D DOA in the presence of impulsive noise," *Journal of Electronics and Information Technology*, vol. 26, no. 3, pp. 350–356, 2004.

[9] H. Jin and L. Zhong, "DOA estimation in impulsive noise environments using fractional lower order spatial temporal matrix," *Acta Aeronautica Et Astronautica Sinica*, vol. 1, no. 27, pp. 104–108, 2006.

[10] L. Li and Q. Tianshuang, "DOA Estimation in Impulsive Noise Environments Using Fractional Lower Order Spatial Temporal Matrix," *Journal of Electronics & Information Technology*, vol. 38, no. 12, pp. 3189–3196, 2016.

[11] W. Juting and J. Shengli, "Subspace based DOA estimation in impulsive noise environments for MIMO radars[J]. Acta Aeronautica Et Astronautica Sinica," *Acta Aeronautica Et Astronautica Sinica*, vol. 30, no. 4, pp. 1653–1657, 2009.

[12] L. Hongsheng, Z. Ruifeng, D. Yu et al., "DOA estimation method of the beamspace in impulsive noise environment," *Modern defence technology*, vol. 45, no. 3, pp. 1653–1658, 2017.

[13] L. Baobao, Z. Junying, and Y. Xiguo, "Angle estimation for bistatic MIMO radar in impulsive noise environments," *Journal of Xidian University*, vol. 24, no. 4, pp. 182–187, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6397973, 11 pages
https://doi.org/10.1155/2018/6397973

Research Article

A Novel Study to Predict Trends and Policies for Mobile Communication in Multienvironment Regions

Shah Jahan Khattak and Bakhtair Khan

Department of Business Administration, Gomal University, DI Khan, KPK, Pakistan

Correspondence should be addressed to Shah Jahan Khattak; khattak@ieee.org

Received 9 July 2018; Revised 11 October 2018; Accepted 18 October 2018; Published 19 December 2018

Guest Editor: Fawad Zaman

The research community has growing interest in the field of mobile phone applications due to its extensive usage in voice, data, video, and their combinations including data-voice and voice-video. The trends of mobile usage have paramount significance to formulate effective policies, to analyze human behavior, reliable prediction, and planning, for development and prosperity of any region in the globe. In this research, we explore the usage of mobile communication to analyze these trends; for this, we select Khyber Pakhtunkhwa (KPK), Pakistan, a multienvironment based on farmers, students, employees, and businessmen. A field survey is carried out by designing a detailed structural questionnaire for viable collection of data to have superior analysis of different multienvironment based classes. A statistical analysis is performed through hypotheses and chi-square test on a large dataset based on sufficient number of observations collected through primary survey for each class. The survey results are provided in number of graphical and numerical illustrations to predict the mobile usage trends, evidently and effectively, of the people of KPK. This study has practical and significant implications in understanding the behavioral intensions of the individuals making policies about mobile phone usage.

1. Introduction

In today's modern and technological era of science and technology, no one can deny the importance of mobile phones and their usage [1]. Mobile phone has become an indispensable and integral part of our life that was initially considered only a source of communication among masses, but now it has key role in creation of employment opportunities, as well as being vital part of accelerating economic growth of any region [2–5]. The major portion of population are thinking that mobile phone applications not only fulfill their requirements but also mobile phone is the source of polite changes at all levels. [6]. In the last three decades, maximum changes occurred in global economy, and development in communication field has produced effective changes in all sectors. No one can deny the potential of mobile usage in diversified fields of applied science, engineering, and technology; see [7–9] and references cited therein.

The purpose of this research is to find the mobile phone usage and the effect of its usage in multienvironment region of Khyber Pakhtunkhwa (KPK), Pakistan, as it may have different impacts on different classes. The present study is a part of a descriptive research, which is based on the primary survey method and performed statistics to predict the behavior of all four multienvironment classes based on farmers, students, employees, and business community. In this research, we exploit the usage of mobile communication to analyze these trends in KPK. A field survey is conducted by formulation of a detailed structural questionnaire for the collection of data to have superior analysis of different multienvironment based classes. A statistical analysis is performed by the use of hypotheses based chi-square tests on a dataset of the primary survey. The survey results are represented by number of graphical and numerical illustrations to predict the mobile usage trends, evidently and effectively, of the people of KPK. The results show that most of the students from University of Engineering and Technology (UET), Islamia College University, University of Peshawar, and Agriculture University, Peshawar, KPK, Pakistan, have a unanimous thinking that mobile phone has brought positive change. Moreover, cell

phone is essential to increase productivity, and at same time, they considered that their fellows have no knowledge of the usage of cell phone at the place of work. This study has paramount significance in understanding the behavioral intensions of the individuals who make policies about mobile phone usage.

The rest of the paper is organized as follows; in Section 2, a brief overview of mobile usage is narrated; Section 3 is dedicated to design methodology; Section 4 presents the results; in Section 5, conclusions with future research direction are listed.

2. Material and Method

Nowadays, the world becomes global due to new emerging technologies and development. New development in technological sector has brought change in the living of society. This study aims to find the impact of mobile phone advancement. In small interval of time, mobile phones have increased connectivity among people. Mobile phones have provided them with the opportunity to remain connected irrespective of their locations. Mobile phone connectivity is available at all locations all the time. Mobile phone has positively contributed toward the relationship among the society individuals. Cell phones have increased the access of people to all levels. Life became easy in all fields and it has also some negative effect. Mobile phone has also created problem of ethics. The dependency on the usage of mobile phone has increased among people. Mobile phone has produced unwanted relationship among the Youngers [10]. The impact of mobile phone can be positive or negative.

2.1. Cellular Communication Dependency. Previously, people used mobile phones as status symbol, but nowadays people have great relay on mobile phone in their network with the society. Leung Louis [11] said that in America most of the Youngers having age ranging from twelve to seventeen years have a mobile phones and mobile is very popular in their social network. In early days, the cost of mobile phone was very high in Asia and limited people can afford it. With passage of time, the prices decreased [9, 10]. These models were heavy in weight and were difficult to carry, but with development new standards were introduced like MMS, Internet, and connectivity which decreased the cost of mobile phones, now available at low prices [12, 13].

Nickerson C. [14] described the term leisure boredom which means that all individuals are in search of becoming busy and they have a greater dependency on mobile phones. Most of the Youngers are busy in Internet, MMS, and voice communication when they are free. Its mean that they pass their free time in making calls to friend or playing games with usage of mobile phones. Mobile phones may sometimes take the Youngers into great risk not following the standard procedures. The Youngers have greater dependency on mobile phones [12]. The common observation is that when the people are free, they become bored and they are in search of becoming busy, while violating the rules of usage of mobile phones. This approach was adapted by the people who greatly depend on mobile phone for the purpose of security and business while being connected with their families. The positive contribution of mobile phone is to make call during incident or to make entertainment by using mobile phone [14].

Youngers think that mobile phone has positively contributed in society and make life easy by connecting with others irrespective of locations. [1]. According to report that mobile phones are tremendously used during emergency, which has saved large number of people, mobile phone has become an integral part of human life.

Mobile phone has also positive contribution toward travelling and toward access to those being far away from their offices [1]. A report indicates that 70% + of people use mobile phone for business activities while minimum percentage of people have personal activities [12].

Nokia Company carried out a survey in year 2001 from three thousand plus people in all over the world under the age of 40. About 70% + of people said that they use MMS function most of all from their mobile phone. People have common usage of text messing. Mobile phones are used in making friendship and having connectivity [15]. Initially mobile phones were used for business. [1]. In USA, the engineers pointed out that the land line was invented for the purpose of business and military usage [16]. According to information presented in [17], which shows that in Australia the mobile phone users are Nineteen billion +, the mobile phones adapted by Youngers are very costly. According to information reported in [17], the maximum percentage of Youngers, having age ranging from ten to sixteen years old, and minimum percentage of adults have a personal cell phone. From a survey of [18] globe cellular penetration is up to 3.7 billion and according to estimate most of the population have personal cell phones, which puts about seventy plus percent of world's population in the range of mobile phones [19]. Mobile phones were among the few technologies which spread so rapidly in the world. People use cell phones for different purposes. The same study showed that almost 60 percent of people keep their mobile phones with them all the time wherever they go [16, 17]. In offices the employers expect that people will be self-responsible and will not use cell phones for personal purposes. Employers usually ask employees to make personal phone use at lunch time or tea break and tell their family and friends not to disturb them during work hours except in case of emergency [20].

2.2. Cellular Communication Dependency and Impact. Development in communication sector has brought revolution in all sections of the world. After 1980, mobile phones replace land lines in developed countries and contributed positively in less developed countries. Mobile phones have great impact on the social life of the society. Males and females have equal opportunities of the usage of mobile phones for network in society. Cell phones are the main tools of production in less developed countries.

Most people use mobile phones for the currency rates, market tendency, and access to abroad [20]. Limited people have thought that mobile phones have negative impact when used in duty places. Small group of people think that mobile phones have increased the chances of access, thus

having positive impact [20]. Mobile phone has contributed to the GDP of developed countries and produced millions of pounds. Mobile phones have very large percentage of impact on social sector. Mobile phones give satisfaction by providing connectivity at all locations [14]. The role of mobile phone cannot be denied in all types of business. There is full information sharing among the customers and owners in market. The physical distance is reduced among the suppliers and buyers. Many bosses think that the usage of mobile phone at duty places has negative impact and the production is reduced. The employers usually think to ban the mobile phone during the period of production. This has some negative impact because the workers will not be comfortable which will resultantly reduce the production. The policy makers should recommend procedures based on the nature of business and the vision of the employees [19]. For example, the mobile phones should not be permitted during travelling and working hours. The employees in different sections of the company especially in marketing and sales should not be allowed to use mobile phones during working hours. This will result in ignoring the customers. Mobile phones should be silent during meetings and conferences; otherwise, they will create negative impact. Depending on the nature of work, supervision level people should be allowed to coordinate with customers and their coworkers. People at reception desk have different nature of work; they may be allowed to remain in contact with colleagues. Mobile phone has positively contributed toward economic growth through, for example, easy access to market, which reduced the cost of communication and also reduced the travelling cost for physical and face to face meeting.

Mobile phone has different roles in different countries. In Europe, it is used to bring unification and rebuild the economy, while in Finland, it is used as status symbol. The mobile phone in Japan is a fast way of communication, while in Philippine, it is the source of political change.

The mobile phone has reduced the travelling of those people involved in much travelling to reach offices and time limitation is there.

The impact of mobile phone is usually negative including reduction in production and disturbances during working hours based on the receiving of calls and the ring tunes of the mobile phone. For a limited time, the concentration of the worker changes whether he receives a call or ignores it. Mobile phone has a negative effect on the workers if they are in production section. In developed countries, million and billions of dollars have been lost due the usage of mobile phones at the duty places. Usually the workers lose their concentration and they are unable to meet the targets. The usage of mobile phones has both positive and negative effects on students. In case of research, it has positive impact, while during class it has negative impact. The best way is to turn off mobile phones during working hours or during class in case of students, this will create positive image of mobile phones.

3. Methodology of Research

Collected data is primary and qualitative in nature. Data was collected from the students of UET Peshawar, Islamia College University, University of Peshawar, and Agriculture University Peshawar; farmers; business community; and employees; all are from KPK, Pakistan. Structured questionnaire has been used as research tool and was distributed among the different classes. The primary data has been collected from the students of UET Peshawar, Islamia College University Peshawar, University of Peshawar, and Agriculture University Peshawar, KPK, for the class of students. The primary data of employees has been collected from various employees working in public sector as well as in private sector. The primary data of farmers has been collected from the farmers of district Karak, KPK. The primary data of business community has been collected from businessmen belonging to Peshawar and from the surrounding areas. The multienvironment means that various classes were selected from the population of KPK. Total enrolment of the people is 100 and 156 belonging to KPK, Pakistan. Few questions were asked by people and they were told to return to the questioner. Different frequencies and suitable percentage were used to explore the collected data.

Mobile phones usage by people, KPK, Pakistan, at duty sites and their impact on people will indicate dependency of the classes on mobile phone.

The overall study has followed the framework presented above, while the workflow diagram is presented in Figure 1.

3.1. Sample and Population. Based on the data collected from people in KPK, Pakistan, for this research study and to analyze collected data, students were selected from UET Peshawar, Islamia College University Peshawar, University of Peshawar, and Agriculture University of Peshawar, KPK, which include students from Electrical Engineering and Computer Engineering, because they have some basic knowledge of mobile communication, while other disciplines have their own domain in KPK. The primary data of farmers in KPK has been collected from the farmers of district Karak, KPK, and business community in KPK, Pakistan. Data was collected from 100 people initially and later on from 156 people including students from UET Peshawar 32%, farmers 23%, business community 25%, and employees 20%. For presentation of the mobile phone applications, data is based on 32 students, 23 farmers, 25 businessmen, and 20 employees.

3.2. Observations Collection Tools. Primary data was collected through structured questionnaire from the students of UET Peshawar, Islamia College University Peshawar, University of Peshawar, and Agriculture University of Peshawar; farmers; employees; and business community, KPK, Pakistan. For data analysis, Excel was used. Observations collected through this tool have been analyzed by using suitable frequencies and percentages.

4. Results and Discussion

Primary data is collected through structured questionnaire from the students of UET Peshawar, farmers, employees, and business community, KPK, Pakistan. For data analysis, Excel was used. Observations collected through this tool have been analyzed by using suitable frequencies and percentages.

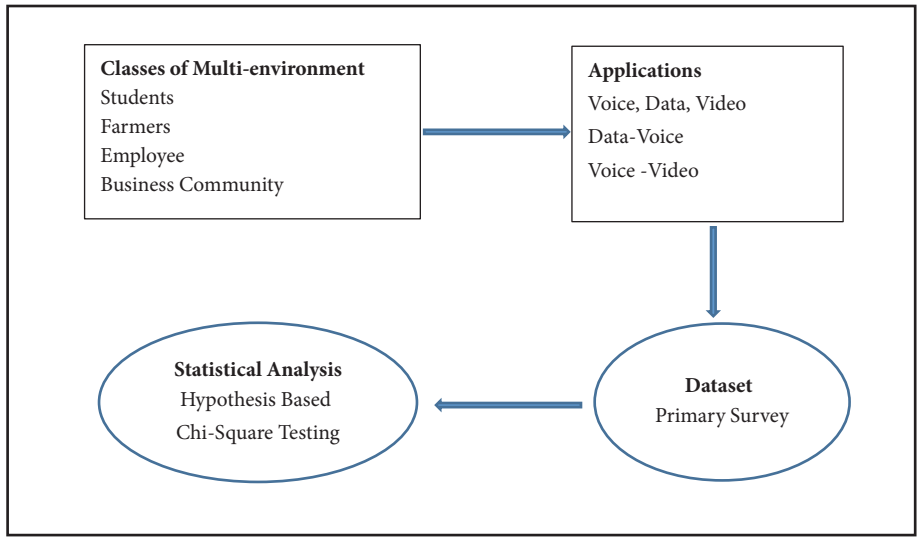


Figure 1: Overview of research model for mobile communication study.

Table 1: Statistical performance indices for all four classes on the basis of 100 samples.

Application	Multi-Environment Classes				Statistical Indices		
	Students	Farmers	employees	Business	Mean	Variance	Standard Deviation
Voice	4	22	7	13	2.63	7.89	2.81
Data	3	1	4	2	2.50	7.50	2.74
Data-Voice	20	0	5	5	1.83	4.83	2.20
Video	3	0	4	2	2.56	7.89	2.81
Voice-Video	2	0	0	3	2.80	10.00	3.16

Table 2: Statistical Performance indices for all four classes on the basis of dataset of 156 samples.

Application	Multi-Environment Classes				Statistical indices			
	Students	Farmers	Employees	Business	SUM	Mean	Variance	Standard Deviation
Voice	12	21	5	14	52	13	32.5	43.33
Data	18	0	6	3	27	6.75	46.68	6.83
Data-Voice	28	0	8	5	41	10.25	113.18	10.64
Video	1	9	2	2	14	3.5	10.25	3.2
Voice-Video	17	0	1	1	19	4.75	50.19	7.08
SUM	76	30	25	25				

The structured questioners consist of five variables, i.e., voice, data, data-voice, video, and voice-video. Based on the average response from the collected observations from farmers, business community, employees, and students of UET Peshawar, KPK, Pakistan, results are summarized in Tables 1 and 2 for dataset based on 100 samples and 156 samples, respectively, while graphical illustrations are presented in Figures 2 and 3 based on classes and application, respectively, for dataset based on 100 samples. Additionally, the graphical illustration results are presented in Figures 4–11 for dataset based on 156 samples in terms of pie graphs and cluster charts.

It is seen from the results presented for 100 samples based dataset that the high value of mean, i.e., 2.80, shows that the people use cell phones mostly for voice-video, which predicts the increased communication among peoples, as well as their behavior toward mobile phones. The 2nd value of mean, i.e., 2.63, for the voice shows that people use cell phones for voice, which indicates the positive attitude toward cell phone usage. It means that that people are well connected. The 3rd value of the mean, i.e., 2.56, for the application of video shows that people use mobile phone for enjoyment and entertainment which may be harmful during working hours. The 4th value of mean, i.e., 2.50, shows that people use mobile phones for data such as downloading data for research, literature, and stories. The 5th value of the mean, i.e., 1.83, shows that people use cell phones for data-voice, which is at the lowest level, and thus one may say that most people have no smart phones for data.

It is seen from the result presented for 156 samples based dataset that the high value of mean, i.e., 13, shows that people

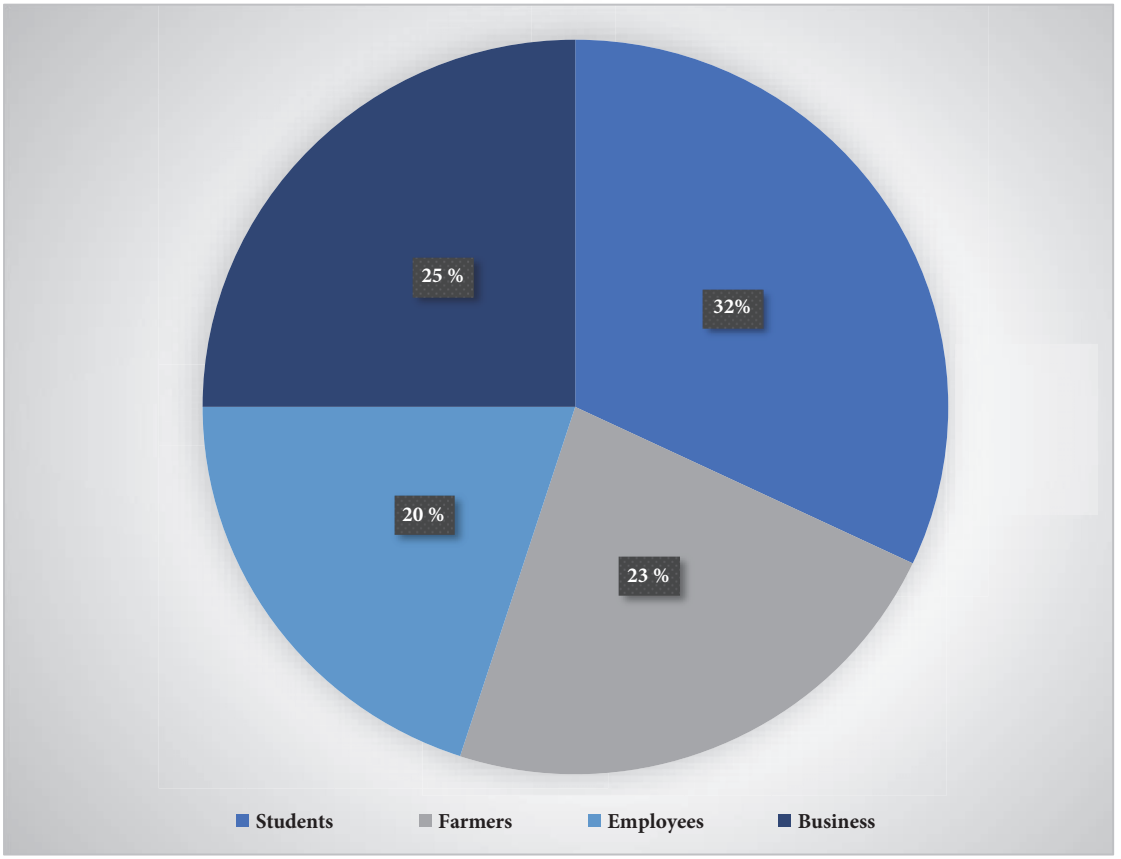


Figure 2: Pie-Chart based graphical illustration of the results for each class based on dataset of 100 samples.

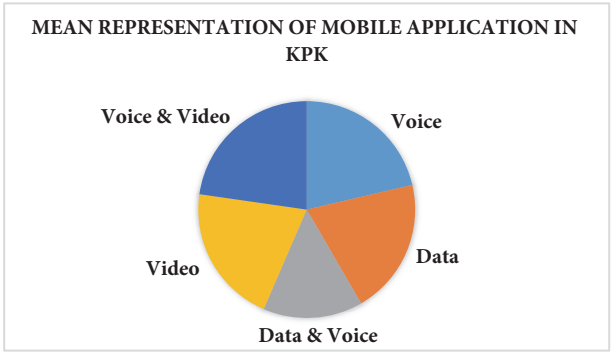


Figure 3: Pie-Chart based graphical illustration of the results for each application based on dataset of 100 samples.

use cell phones mostly for voice, which predicts the increased communication among people, as well as their behavior toward mobile phone. The 2nd value of mean, i.e., 10.25, for the data-voice shows that people use cell phone for voice as well as data, which indicates the positive attitude toward cell phone usage. It means that people are well connected. The 3rd value of the mean, i.e., 6.75, for the application of data shows that the people use mobile phones for enjoyment and entertainment, which may be harmful during working hours. The 4th value of mean, i.e., 4.75, shows that the people use mobile phones for data such as entertainment and voice calling. The 5th value of the mean, i.e., 3.5, shows that people use cell phone for only video, which is at the lowest level, and thus one may say that most peoples have no smart phones for only video.

5. Conclusion

Based on the new technologies and new feature, mobile phone became integral part of our life. At work places, people have serious concern about mobile phone. Side SMS by student during class timing can waste the time of the students. This research has been carried out to predict the trend of mobile phone application among people of Pakistan. Observations were collected by questioner from students of UET Peshawar, Islamia College University Peshawar, University of Peshawar, and Agriculture University of Peshawar; farmers; business community; and employees. Analysis indicates that most students use mobile phones for voice conversion with family and coworkers. Mobile phones can be used for entertainment and text messaging. The result tabulated indicates that most people claim that they are not disturbed due to mobile phone. However, they would prefer to turn off their mobile phone during working hours. Most students receive around ten calls in average, which is comparatively high number. Based on the analysis from the observations

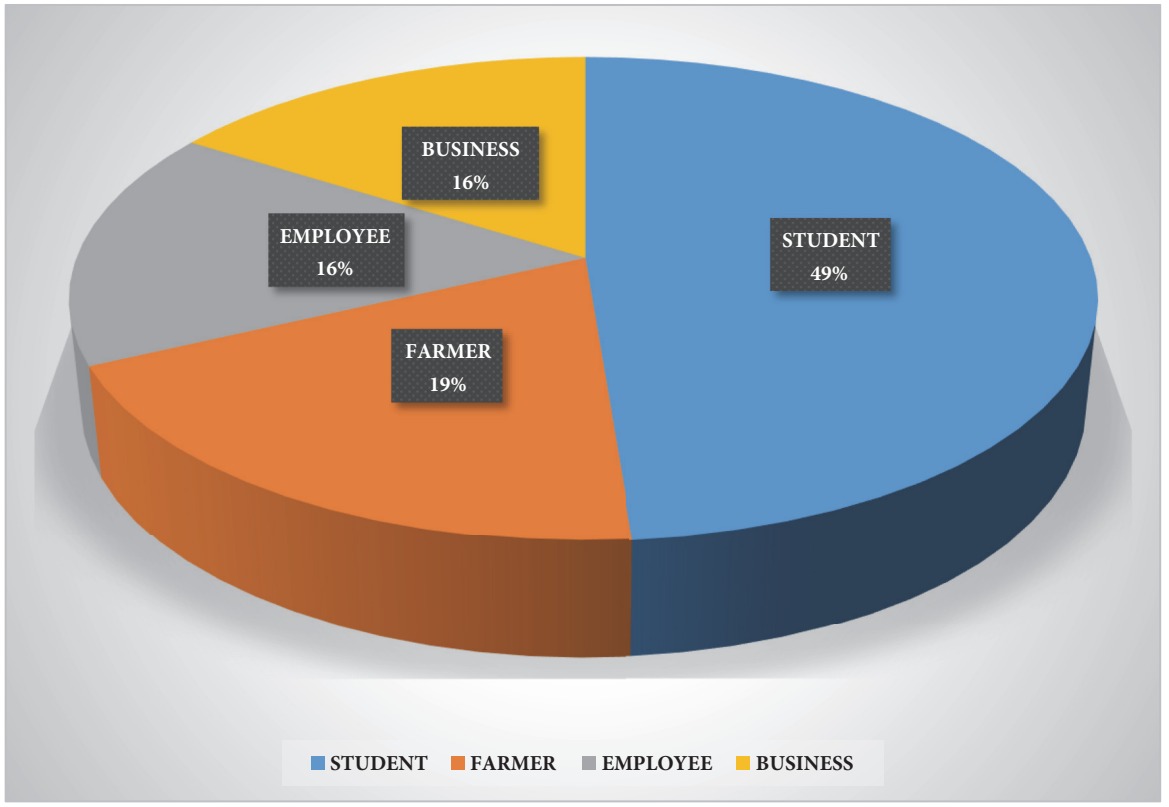


Figure 4: Pie-Chart based graphical illustration of the results for each class based on dataset of 156 samples.

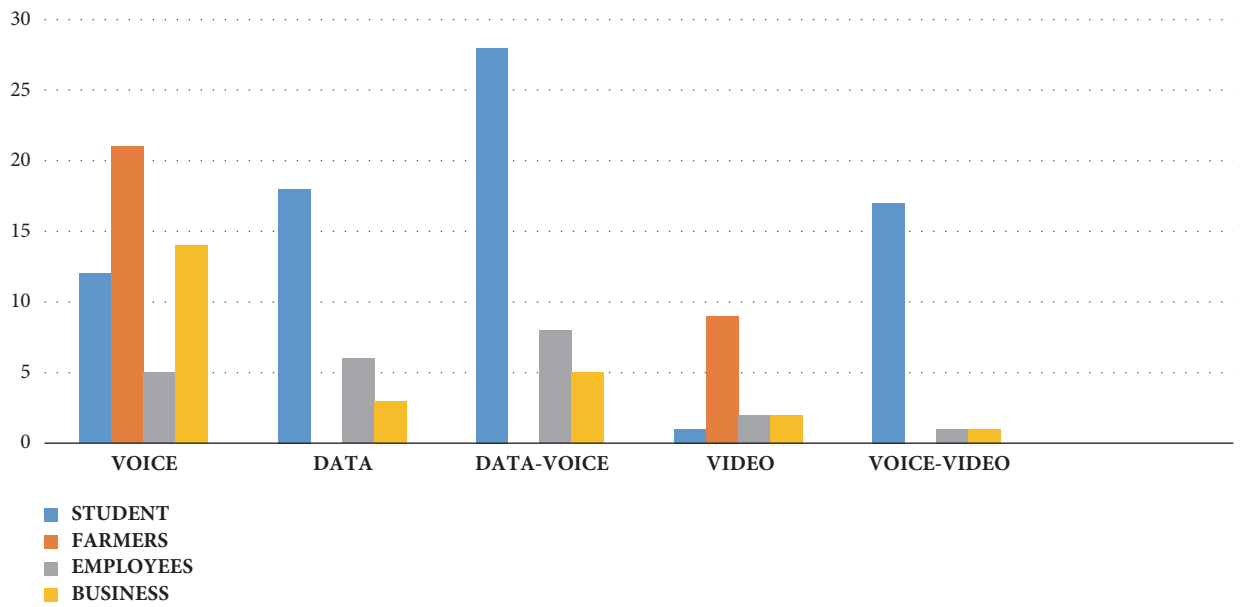


Figure 5: Cluster-Chart based graphical illustration of the results for each application based on dataset of 156 samples.

collected, over 50+ percent of these calls are social in nature. These calls from mobile phones have increased the efficiency among the friends and families. Most of the mobile phones are not used for data.

A large number of people have the view that most of their work is related to mobile phones and their productivity is highly based on the usage of mobile phones, which shows the positive impact of mobile phones.

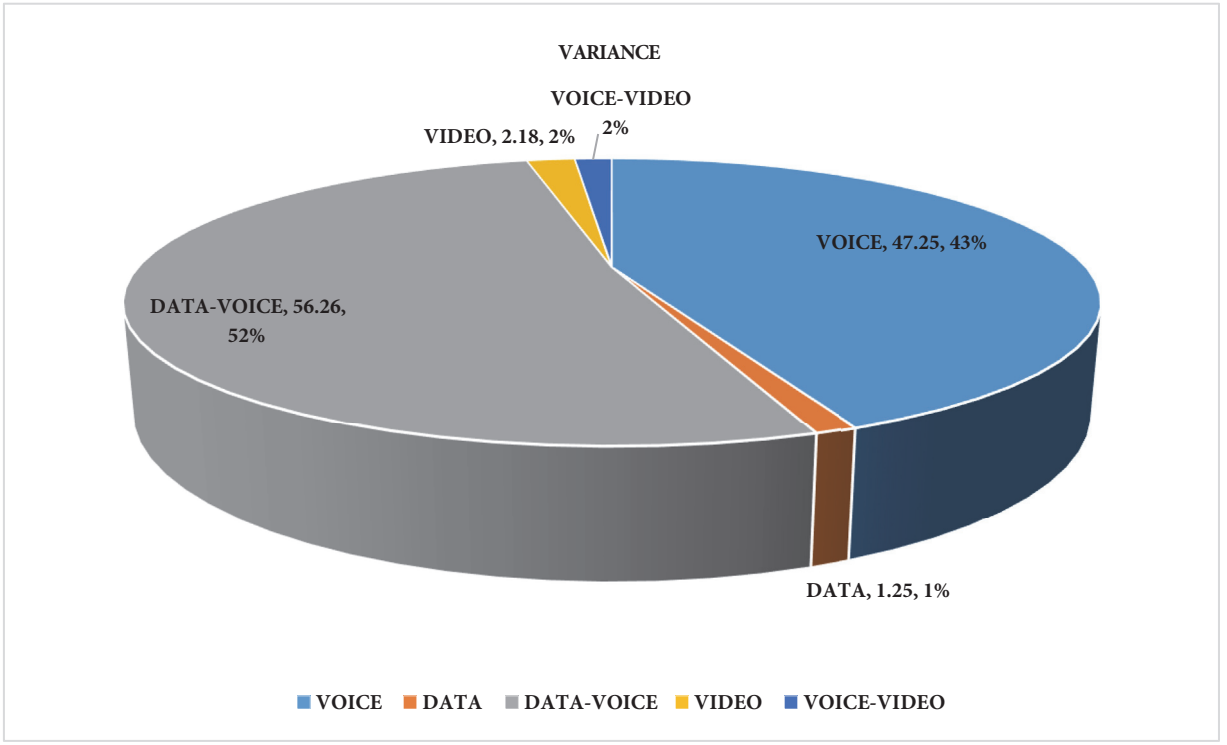


Figure 6: Pie-Chart based graphical illustration of the results for each application based on dataset of 156 samples.

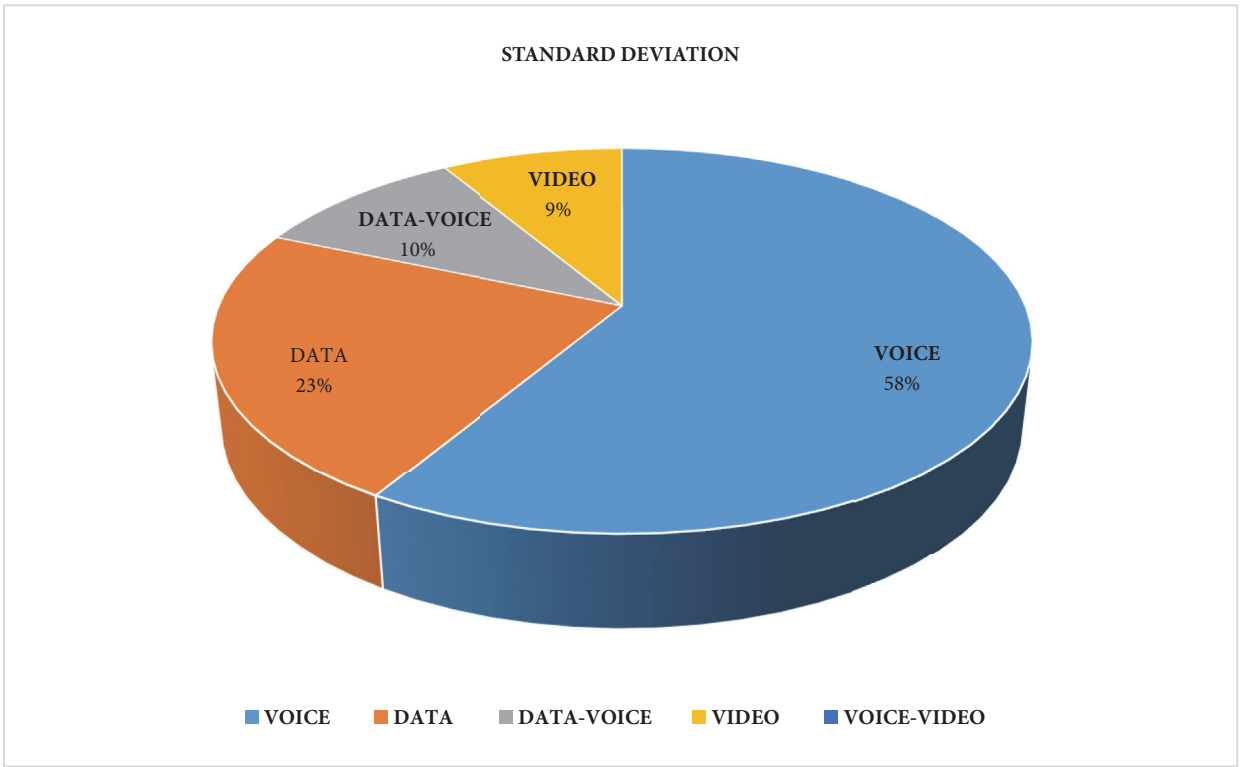


Figure 7: Pie-Chart based graphical illustration of the results for each application based on dataset of 156 samples.

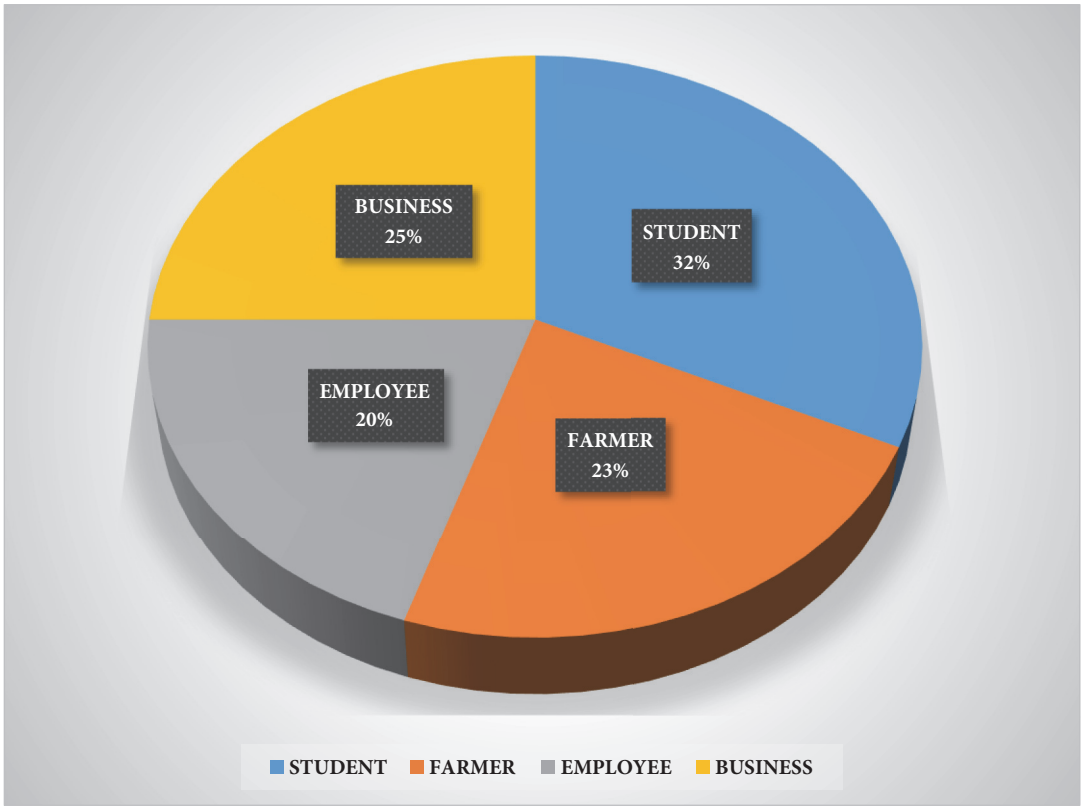


Figure 8: Pie-Chart based graphical illustration of the results for each class based on dataset of 156 samples.

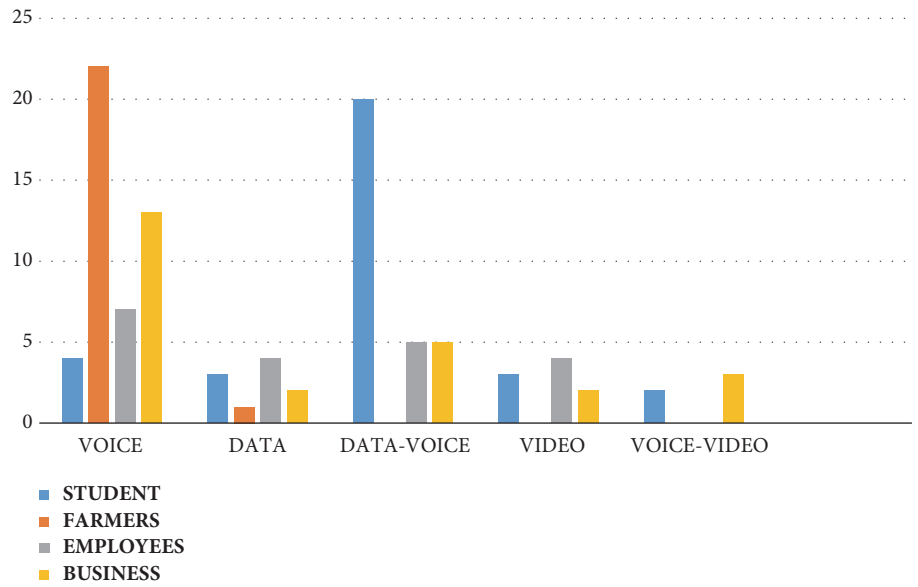


Figure 9: Cluster-Chart based graphical illustration of the results for each application based on dataset of 156 samples.

The report based on the observations collected from the students indicates that the use of mobile phone has increased in all sectors of society including students and other classes of society. Dependency of mobile phones has increased in institutions like universities, companies, and other sectors of production. Mobile phone has made access to information easy especially in jobs related to information. The employers and the owners use the mobile phone for market information. Physical distance has decreased and the people are well connected.

One group of people considered that they cannot live without the mobile phones and they do not violate the

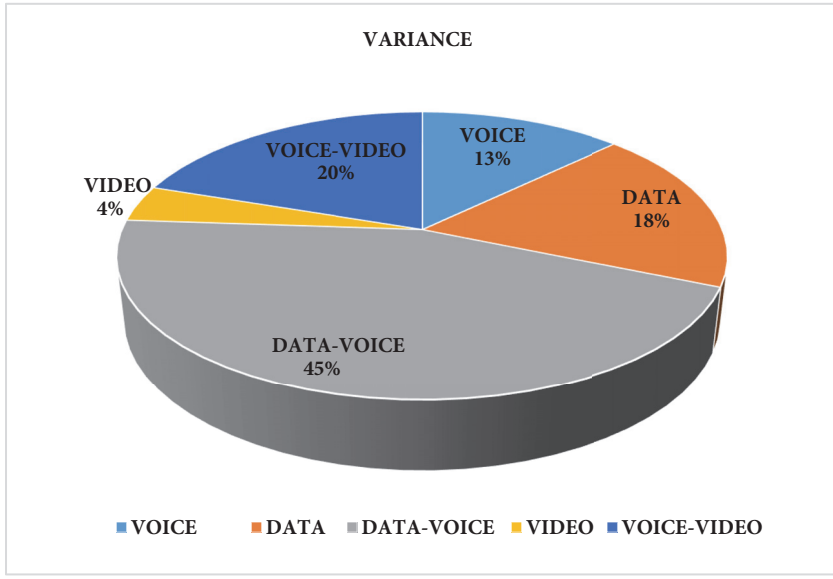


Figure 10: Pie-Chart based graphical illustration of the results for each application based on dataset of 156 samples.

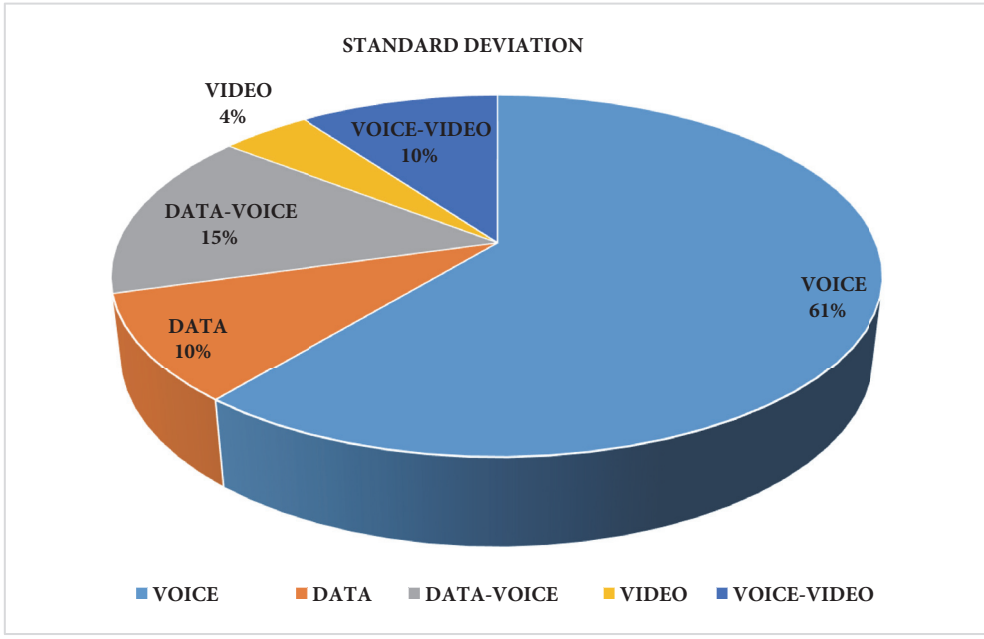


Figure 11: Pie-Chart based graphical illustration of the results for each application based on dataset of 156 samples.

standard procedures with ultimately increased productivity. The decisions based on these observations show that mobile phones can be used more productively.

At higher levels of management, the use of mobile phones is considered as positive change for business and corporate sectors. The mobile phones have developed good mutual relation of trust among the business community. A large number of people have the opinion that the mobile phone has brought a positive change in all fields of life. This research was based on the primary data collection, which was effective and result oriented, and it predict trends and policy for users from different classes.

In future, one may investigate neural networks [21–26] based methodologies for better planning, prediction, and policies for not only KPK region but also other similar regions all over the world.

Appendix

Questionnaire

Techno-Economic Impact of Mobile Communication on Society in Context of KPK, Pakistan. This questionnaire is developed for the survey of my Ph.D. research. The basic aim of this

Table 3: Demand side.

S.No	Description	5	4	3	2
1	Age Group	18-28	29-38	39-48	49-68
2	Working Class	Student	Farmer	Employee	Business
3	Province	KPK	Sind	Punjab	Baluchistan
4	Location of Living	Urban	Suburban	Rural	Sub Rural
5	Demand	Extreme	Needed	Normal	No Demand
6	Quality of Demand	Excellent	Better	Good	No Quality

Table 4: Supply side.

S.No	Description	5	4	3	2
1	Quality	Excellent	Better	Good	No Quality
2	Cost	Very High	High	Normal	Cheap
3	Satisfaction	Excellent	Better	Good	Normal
4	Living	Excellent	Better	Good	Normal
5	Willing to pay	Yes	Up to some Extent	No	Not at all
6	Employment opportunities	Yes	Upto some extent	No	Not at all
7	Impact	Positive	Negative	Positive & Negative	No. Impact
8	Revenue Contribution	Yes	Upto Some extent	No	Not at all
9	Need	Yes	No	Yes & No	Not at all

Table 5: Mobile network side.

S.No.	Description	5	4	3	2	1
1	Mobile operator	Mobil ink	Telenor	Ufone	Warid	Zong
2	Generation of Network	4G	3G	2.5G	2G	1 G
3	Year of User	1981-86	1987- 93	1994-2000	2001-2007	2008- 2014
4	Application	Voice	Data	Data & Voice	Video	Voice & Video

survey is to collect the information of the technological growth of the mobile communication and the economic impact demand side, impact on supply side, impact on GDP, and impact on taxation on society. Your participation is highly needed in this dissertation survey.

Thanks. Shah Jahan Khattak Ph.D. Scholar
Gomal University, Dera Ismail khan/CESET, Islamabad.

1 Name ----------
2 Gender Male Female
3 Marital Status Single Married Widow Divorced

See Tables 3, 4, and 5.

Data Availability

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

Conflicts of Interest

All the authors of the manuscript declared that there are no potential conflicts of interest.

References

[1] M. Castells, M. Fernandez-Ardevol, J. L. Qiu, and A. Sey, "Mobile Communication and Society: A Global Perspective," *Economic Geography*, vol. 84, no. 1, pp. 119-120, 2008.

[2] M. Myint, A. Adam, S. Herath, and G. Smith, "Mobile phone applications in management of enuresis: The good, the bad, and the unreliable!," *Journal of Pediatric Urology*, vol. 12, no. 2, pp. 112–112.e6, 2016.

[3] R. Whittaker, R. Borland, C. Bullen, R. Lin, H. McRobbie, and A. Rodgers, "Mobile phone-based interventions for smoking cessation," *Cochrane Database of Systematic Reviews*, no. 3, 2007.

[4] L. F. Garabedian, D. Ross-Degnan, and J. F. Wharam, "Mobile Phone and Smartphone Technologies for Diabetes Care and Self-Management," *Current Diabetes Reports*, vol. 15, no. 12, article no. 109, 2015.

[5] M. Stanley, Y. Huang, T. Loh, Q. Xu, H. Wang, and H. Zhou, "A high gain steerable millimeter-wave antenna array for 5G smartphone applications," in *Proceedings of the 11th European Conference on Antennas and Propagation, EUCAP 2017*, pp. 1311–1314, March 2017.

[6] Y. Yuan, M. Raubal, and Y. Liu, "Correlating mobile phone usage and travel behavior - A case study of Harbin, China," *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 118–130, 2012.

[7] A. Backholm, M. Tervahauta, S. Salorinne et al., "Seven Networks Inc," in *Dynamic adjustment of keep-alive messages for efficient battery usage in a mobile network*, pp. 15–883, U.S. Patent Application, 2018.

[8] A. Orsino, A. Ometov, G. Fodor et al., "Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical MTC in 5G," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 79–87, 2017.

[9] P. Rost, C. Mannweiler, D. S. Michalopoulos et al., "Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[10] A. Lepp, J. E. Barkley, and A. C. Karpinski, "The relationship between cell phone use, academic performance, anxiety, and Satisfaction with Life in college students," *Computers in Human Behavior*, vol. 31, no. 1, pp. 343–350, 2014.

[11] L. Leung, "Leisure boredom, sensation seeking, self-esteem, and addiction: Symptoms and patterns of cell phone use," in *Mediated Personal Communication*, S. B. Barnes, E. A. Konijn, M. Tanis, and S. Utz, Eds., pp. 359–381, Routledge, New York, NY, USA, 2008.

[12] R. K. Donya, "Cellular Phone Usage and Productivity among Employees in A Ghanaian SME: An Assessment," *International Journal of Computing and ICT Research*, vol. 5, no. 1, pp. 21–34, 2011.

[13] R. W. Hahn and J. E. Prieger, "The impact of driver cell phone use on accidents," *The BE Journal of Economic Analysis & Policy*, vol. 6, no. 1, 2006.

[14] R. C. Nickerson, H. Isaac, and B. Mak, "A multi-national study of attitudes about mobile phone use in social settings," *International Journal of Mobile Communications*, vol. 6, no. 5, pp. 541–563, 2008.

[15] D. J. Reid and F. J. M. Reid, "Text or talk? Social anxiety, loneliness, and divergent preferences for cell phone use," *Cyberpsychology, Behavior, and Social Networking*, vol. 10, no. 3, pp. 424–435, 2007.

[16] C. Fisher, "Cardiologists embrace new terminology," *BMJ*, vol. 330, no. 7501, p. 1207, 2005.

[17] N. Cohen, "What Works: Grameen Telecom's Village Phones? A Digital dividend Study By The World Resources Institute," in *World Resources Institute–Digital Dividend*, 2001.

[18] B. Schackner, "More students opt for wireless phone: colleges face loss of commissions," in *Knight Rider Tribune Business News*, 2002.

[19] J. Chan, Z. Chen, I. Cormane, N. Her, and R. Thomas, "Cell phone industry analysis," in *The multidimensional benefit of mobile technology in the developing world*, Diego R., Ed., ICT for Development Community, 2006.

[20] M. N. Baliamoune, "The new economy and developing countries: Assessing the role of ICT diffusion," WIDER Discussion Papers//World Institute for Development Economics (UNU-WIDER) 2002/77, 2002.

[21] Y. Wang, R. Li, H. Zhang, H. Tan, and Q. Chai, "Using Sentence-Level Neural Network Models for Multiple-Choice Reading Comprehension Tasks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2678976, 8 pages, 2018.

[22] B. Tang, J. Hu, X. Wang, and Q. Chen, "Recognizing Continuous and Discontinuous Adverse Drug Reaction Mentions from Social Media Using LSTM-CRF," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2379208, 8 pages, 2018.

[23] G. Q. Yang, "Multilayer Learning Network for Modulation Classification Assisted with Frequency Offset Cancellation in Satellite to Ground Link," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 1372439, 13 pages, 2018.

[24] M. Kanaan and M. Suveren, "In-body ranging with ultra-wideband signals: Techniques and modeling of the ranging error," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 4313748, 15 pages, 2017.

[25] Q. Zhao, S. Lyu, B. Zhang, and W. Feng, "Multiactivation Pooling Method in Convolutional Neural Networks for Image Recognition," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8196906, 15 pages, 2018.

[26] M. A. Raja, A. Mehmood, A. u. Rehman, A. Khan, and A. Zameer, "Bio-inspired computational heuristics for Sisko fluid flow and heat transfer models," *Applied Soft Computing*, vol. 71, pp. 622–648, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8349486, 6 pages
https://doi.org/10.1155/2018/8349486

Research Article

Flexible Queuing Model for Number of Active Users in Cognitive Radio Network Environment

Ahsan Tanveer[1], **Z. U. Khan**[2], **A. N. Malik**[2] **and I. M. Qureshi**[3]

[1]*Department of Computer Sciences, Iqra University Islamabad, Pakistan*
[2]*Department of Electronic Engineering, International Islamic University, Islamabad, Pakistan*
[3]*Department of Electrical Engineering, Air University, Islamabad, Pakistan*

Correspondence should be addressed to Ahsan Tanveer; hafizahsan_pk@yahoo.com

Received 3 September 2018; Revised 18 November 2018; Accepted 24 November 2018; Published 19 December 2018

Guest Editor: Sungchang Lee

This work presents a Soft Queuing Model (SQM) for number of active users present in a cognitive radio network (CRN) at some given instant. Starting with the existing cellular network where the upper limit for the number of channels and active users is well defined. The idea is then extended to the complicated scenario of CRNs where the upper limit is not deterministic for both the number of channels and the active users. Accordingly a probabilistic SQM is proposed under the condition that the number of channels and active cognitive users are both random variables. The proposed model will be useful to offer the level of reliability to the clients connected with CRN and hence to offer secure communication even on a cooperative CRN. The proposed model has been verified theoretically and simulations have been carried out in diversified set ups to evaluate the performance.

1. Introduction

Cognitive Radios (CR) seems to be the prominent area of research in future communication systems including 5G which is based on Massive MIMO and heterogeneous networks [1, 2]. A reliable communication model for cognitive radio is the requirement of the time. With the development of the idea, its implication in almost all the existing communication setups, including 4G and mobile communication, and the momentum it has gained, it is definite to continue its research importance in future as well [3, 4]. Recent works on smart grid for future energy requirements also incorporate the idea of cognitive radio for communication in different layers of networks; i.e., the network of home appliances, the community, and then the mega scale networks up to metropolitan level are supposed to be exploiting the idea of cognitive radio for communications [5, 6].

Surprisingly, a reliable channel model is still missing for cognitive radio. Specifically if we get the idea of cooperative users with a fusion centre operating in community in parallel with the primary networks and service providers, then, how much resources will be available to cognitive radio network is a basic question. This will also link with the level of reliability that a cognitive radio network may offer to its clients.

There have been lot of research in CR, however, that focus on three major areas. The first one is the channel sensing [7, 8] while second one is the access of spectrum holes or channels [9] and the third one is optimization of resources to utilize the available channels for maximum throughout [10].

Now in case of the reliable communication on cognitive radio network, the question will be how many users are present at some given instant [11] and how many channels are available to accommodate these users [12]? Answer to the questions lie in probability theory. The available channels to cognitive radio network will be based on the parameters such as the overall available/free channels in a primary base station operating in the region and of course on the sensing capability of the cognitive radio network under consideration.

The second question related to the above discussed three major areas of CRN will have certain channels available and will be offering service to its clients; then, how many clients can be accommodated at a given instant is a question whose answer lies in queuing theory.

A queuing model for number of channels occupied by the users in primary radio network is relatively less complicated and has been derived similar as in [13]. However, that is for infinite resources, which has been modified for the limited available resources; i.e., maximum available channels are assumed to be finite. That seems logical and is being employed in existing primary radio networks.

In third phase the idea has been extended to the cognitive radio networks with flexible upper bound. Since the upper limit for available channels in cognitive radio networks is flexible hence, accordingly, a Flexible Queuing Model is being proposed in this paper. The model has been implemented in simulations and the results for diversified scenarios have been presented.

Rest of the paper has been organized as follows: Section 2 presents mathematical framework and proposed model for existing networks with fixed upper limit for number of users. In Section 3 proposed Flexible Queuing Model for cognitive radio networks is discussed. Section 4 presents the whole system model. Simulation results are given in Section 5 and the conclusion is included in Section 6.

In the document p_j represents probability of j number of active users, while **n** and **k** are random variables representing, respectively, the available number of channels and number of active users for cognitive radio network.

2. Mathematical Framework and Proposed Model for Existing Networks

In case of existing mobile network models, the BS dedicates a fraction of available communication channels to accommodate the smooth hand over (HO). The HO is predictable based on the physical location of Mobile user; i.e., the handover probability will be on higher side near the cell boundaries. Yet there could be other causes for handover such as Quality of Service (QoS) and Conjunction; however, they are secondary reasons in comparison.

In case of cognitive radios the available number of channels and hence the active users are random, depending on the sensing capabilities and occupying the available free spots for the spectrum. Hence the Flexible Queuing Model was the ultimate requirement to address this issue. Since the model was not available in the existing literature, therefore this model has been proposed to deal with two different scenarios which may be considered as an extension of M/M/1.

Scenario-I. This case is considered in current Section 2 where the length of queue has been taken as finite. Yet this length is kept fixed. Accordingly the mathematical model has been developed and results have been derived for finite number of users (K in the manuscript).

Scenario-II. The major contribution in our work is the environment in which the finite number K is taken as a random. Accordingly the phrase "Flexible Queuing Model" has been introduced in the title. All the derivations have been carried out and results have been generated for a stochastic upper limit as in Section 3.

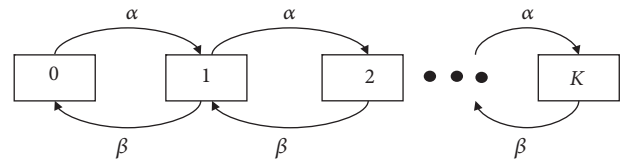


Figure 1: Channel occupation and channel release model for active users in existing networks.

In case of existing cellular networks, the number of available channels for a BS is considered to be deterministic and fixed. Let this number be N. Moreover let the maximum allowable number of active users be K. Therefore, the extra channels, i.e., $h_{extra} = N - K$, will be reserved for handovers. The ratio of maximum allowable active users to the available channels reflects the worst situation that appears in peak hours when the entire K allowable channels will be occupied. As a matter of routine the number of active users will be random occupying a value from the set $\{0, 1, ...K\}$. Assuming k number of active users at an instant when another user appears, the number will move to $k + 1$ state and conversely; if an active user leaves the network, the system will move to $k - 1$ state. Ultimately the number of active users will follow a queuing system model as illustrated in the Figure 1.

In the figure α is the rate at which a new user appears and the process moves forward from state k to $k + 1$ and $0 \le k \le K - 1$ and β is the rate at which process moves backward from state k to $k - 1$ when an active user leaves the system with the condition $1 \le k \le K$.

Accordingly the probability of finding the process in state k at some given instant will be

$$p_k = \varsigma^k p_0, \quad 0 \le \mathrm{k} \le K \tag{1}$$

where $\varsigma = \alpha/\beta$ and p_k is the probability of finding the process in state k at some given instant. The detailed derivation may follow through the queuing model presented in [13].

In order to compute p_0, i.e., the probability of finding the process in state-o at some given instant, we may apply the normalization i.e. $\sum_{k=0}^{K} p_k = 1$

That is, $(1 + \varsigma + \varsigma^2 + \ldots + \varsigma^K)p_0 = 1$.

Now the series, $(1 + \varsigma + \varsigma^2 + \ldots + \varsigma^K) = (1 + \varsigma + \varsigma^2 + \ldots) - (\varsigma^{K+1} + \varsigma^{K+2} + \ldots)$, which is the difference of two infinite series hence will be given as

$$\begin{aligned} p_0 \sum_{k=1}^{K} \varsigma^k &= p_0 \sum_{k=0}^{\infty} \varsigma^k - p_0 \varsigma^{K+1} \sum_{k=0}^{\infty} \varsigma^k \\ p_0 \sum_{k=0}^{\infty} \varsigma^k - p_0 \varsigma^{K+1} \sum_{k=0}^{\infty} \varsigma^k &= p_0 \left(1 - \varsigma^{K+1}\right) \sum_{k=0}^{\infty} \varsigma^k = 1 \end{aligned} \tag{2}$$

It will converge provided $\varsigma < 1$ and hence

$$p_0 \left(1 - \varsigma^{K+1}\right) \frac{1}{1 - \varsigma} = 1 \tag{3}$$

This gives

$$p_0 = \frac{1 - \varsigma}{1 - \varsigma^{K+1}} \tag{4}$$

Thus

$$p_j = \varsigma^j \frac{1-\varsigma}{1-\varsigma^{K+1}} \tag{5}$$

with the condition that $\varsigma < 1$ and hence $\alpha < \beta$. Therefore in this case the forward transition rate α which is the rate at which some new channel is needed must be less than the reverse transition rate β, which is the rate at which the channel is released by the user and is added back in the queue of available resources, which will be the ultimate condition for global balance.

These transition probabilities and queuing model are valid for the existing cellular networks. Below is a little bit complicated model for cognitive radio network. This model will be applicable to the cognitive radio networks committed to provide specific level of reliability to the users.

3. Proposed Flexible Queuing Model for Cognitive Radio Networks

In above section we have proposed a queuing model for cellular networks. In this case a state is an indicator of the active number of users at some given instant. The advantage in case of above derivation was that the maximum allowable active users and number of channels, i.e., K and N, were known to us and HO occurs with relatively less probability that was only p_j. However, in case of CRN the situation will be little bit more complicated; i.e., the available number of channels in this case will be a random variable $\mathbf{n}$ which may take a value from the set $\{0, 1, 2, \ldots, N\}$ with respective probabilities changing with outcome. Evaluation of these probabilities may not be simple, specifically when there is the existence of multiple CRNs operating in parallel on cooperative basis. Yet having $\mathbf{n}$ available channels the CRN will be in position to accommodate $\mathbf{k}$ users which will be decided based on the level of reliability, i.e.,

$$h_{ratio} = \frac{h_{extra}}{h_{active}} = \frac{\mathbf{n}-\mathbf{k}}{\mathbf{k}} \Longrightarrow \mathbf{k} = \frac{\mathbf{n}}{h_{ratio}+1} \tag{6}$$

Ratio in (6) is decided by the CRN fusion centre. Accordingly for $\mathbf{k}$ active users, $\mathbf{m} = h_{ratio}.\mathbf{k}$, where $\mathbf{m} = \mathbf{k} \leq \mathbf{n}$ will be upper bounded, where $\mathbf{n}$ is the total number of channels available to CRN after sensing and it will be random as well.

$$\begin{aligned} \mathbf{k} + \mathbf{m} = \mathbf{n} &\Longrightarrow \\ \mathbf{k} + \mathbf{k}h_{ratio} = \mathbf{n} &\Longrightarrow \\ \mathbf{k} = \frac{\mathbf{n}}{1+h_{ratio}} & \end{aligned} \tag{7}$$

This $\mathbf{k}$ represents the maximum allowable users. Obviously $\mathbf{k}$ will also be a random variable in this case. Depending on this set up, the queuing system, as developed in case of above cellular model, will need to be flexible. The updated Figure 2 is given below.

The transition rates α_c and β_c in this scenario will be greater than their corresponding counterparts in previous section, i.e., $\alpha_c > \alpha$ and $\beta_c > \beta$, which indicates more vibrant environment in this case.

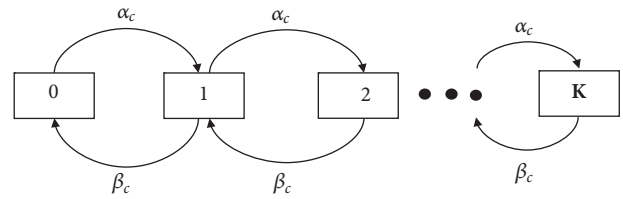


Figure 2: State model for channel occupation and release by active users in cognitive radio network.

Theorem 1. *In given scenario, the state probabilities p_j and condition for convergence are derived as below.*

Proof. Given fixed $\mathbf{k}$, i.e., $\mathbf{k} = K$, the state probabilities p_j will become

$$p_j = \lambda^j \frac{1-\lambda}{1-\lambda^{K+1}}, \quad 0 \leq j \leq K \tag{8}$$

The scenario of expression (8) is similar to that discussed in equation (5). However present scenario is flexible and $\lambda = \alpha_c/\beta_c$. Similarly the condition for convergence will be $\beta_c > \alpha_c$. Since $\mathbf{k}$ is random, the above probability in this case will become

$$p_j = \sum_{k=1}^{K} \frac{\lambda^j (1-\lambda)}{1-\lambda^{k+1}} P(\mathbf{k} = k) \tag{9}$$

where $P(\mathbf{k} = k)$ will be derived from $P(\mathbf{n})$ available channels for CRN which is based on the sensing strategies, the environment temperature, and number of active CRNs in given area. The expression in (9) is proposed with the help of probability theory and specifically total probability theorem.

Rearranging expression (9) we get

$$\begin{aligned} p_j = \lambda^j (1-\lambda) \sum_{k=0}^{K} \frac{P(\mathbf{k}=k)}{1-\lambda^{k+1}} &= \lambda^j (1-\lambda) \\ \cdot \left(\frac{P(\mathbf{k}=0)}{1-\lambda} + \frac{P(\mathbf{k}=1)}{1-\lambda^2} + \ldots + \frac{P(\mathbf{k}=K)}{1-\lambda^{K+1}} \right) & \end{aligned} \tag{10}$$

□

4. Whole System Model

Following flowchart in Figure 3 describes the whole system model including the architecture of primary and cognitive radio user environment.

5. Simulations

In order to verify the performance of proposed models, simulations are carried out in MATLAB. For this purpose different cases, under different situations, are considered.

Case 1. In Case 1, the probability of number of active users for existing cellular networks is determined. For this purpose, the number of available channels and maximum number of

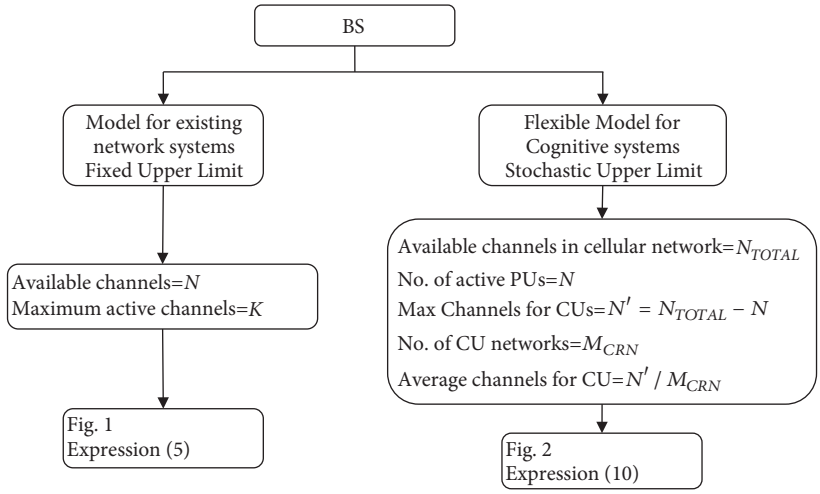


Figure 3: Model for channel occupation and release by active users in existing and CR networks.

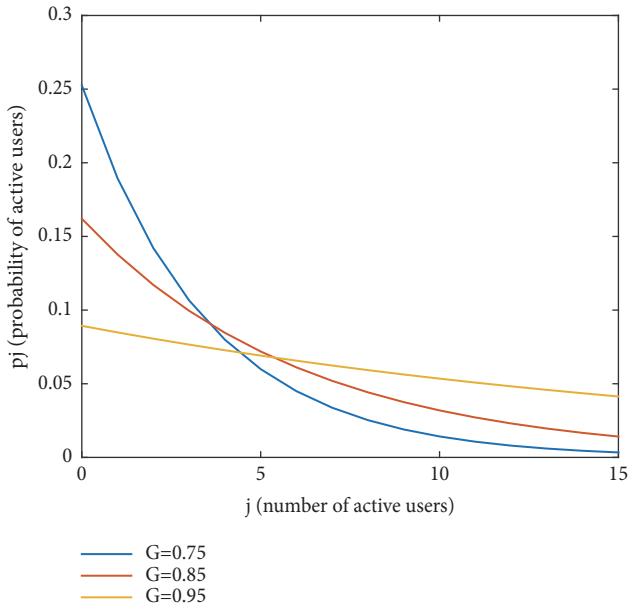


Figure 4: Probabilities of different number of active users for $\varsigma = 0.75, 0.85$ and 0.95.

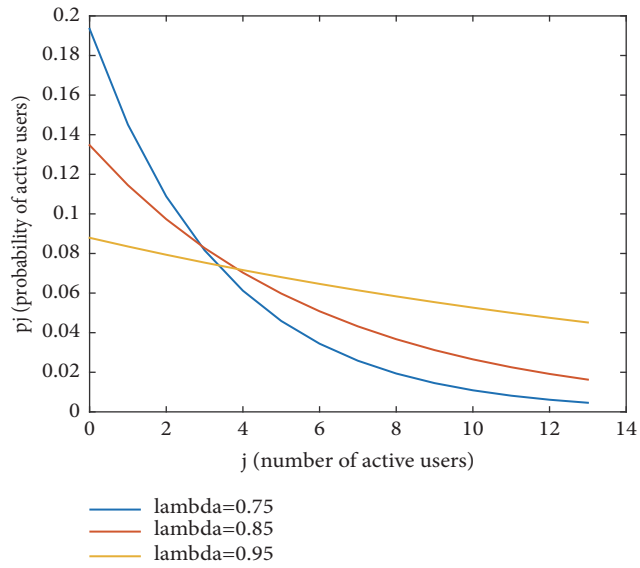


Figure 5: Probabilities of active cognitive users for Case 2 when variance of $\sigma = 3$.

active users for base station are taken as 20 and 15, respectively ($N = 20, K = 15$). Random variable $\mathbf{k}$ representing number of active users will occupy values from the set $\{0, 1, \ldots, 15\}$. Required probabilities are determined using expression (5) for different values of ς, i.e., 0.75, 0.85, and 0.95. The corresponding probabilities are shown in Figure 4.

Case 2. This case deals with probability of active users for cognitive radio network. For this purpose let N = number of active users for primary BS=100.

N' = number of vacant channels for cognitive users $= N/2 = 50$

M_{CRN} = number of cognitive networks = 5

α = average number of channels per cognitive network $= N'/M_{CRN} = 10$

Suppose n = variance of $\sigma = 3$.

The probability of active users for CRN is evaluated using Poisson random variable distribution in expression (10) and the resulting expression becomes

$$p_j = \lambda^j (1-\lambda)\, e^{\alpha} \sum_{k=\alpha-n}^{\alpha+n} \frac{1}{1-\lambda^{k+1}} \frac{\alpha^k}{k!} \tag{11}$$

The resultant probabilities for $\lambda = 0.75, 0.85, 0.95$ are shown in Figure 5.

Case 3. For Case 3 simulations are performed under the same condition as Case 2 except the variance of σ which is taken to be as 5 and the simulation results are shown in Figure 6.

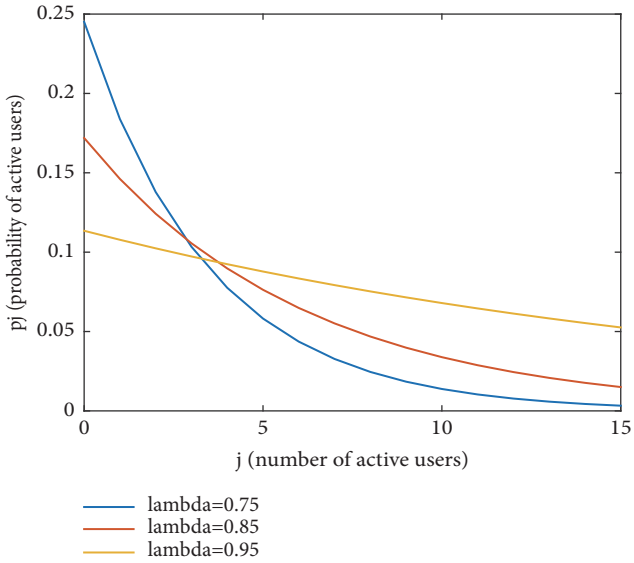


FIGURE 6: Probabilities of active cognitive users for Case 3 when variance of $\sigma = 5$.

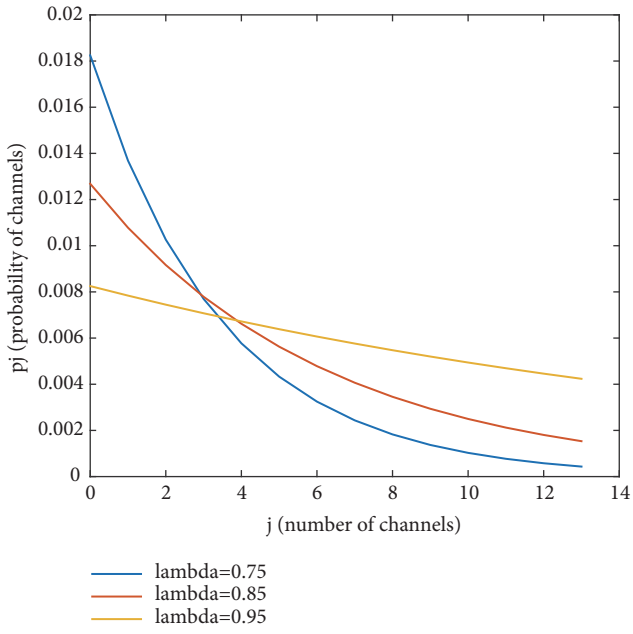


FIGURE 7: Probabilities of active cognitive users for Case 4 when variance of $\sigma = 3$.

Case 4. Case 4 is same as Case 2 except that the probability of active users for CRN is evaluated using uniform random variable distribution in expression (10) and the resulting expression becomes

$$p_j = \lambda^j (1-\lambda) \frac{1}{N+1} \sum_{k=\alpha-n}^{\alpha+n} \frac{1}{1-\lambda^{k+1}} \tag{12}$$

The resultant probabilities for $\lambda = 0.75, 0.85, 0.95$ are shown in Figure 7.

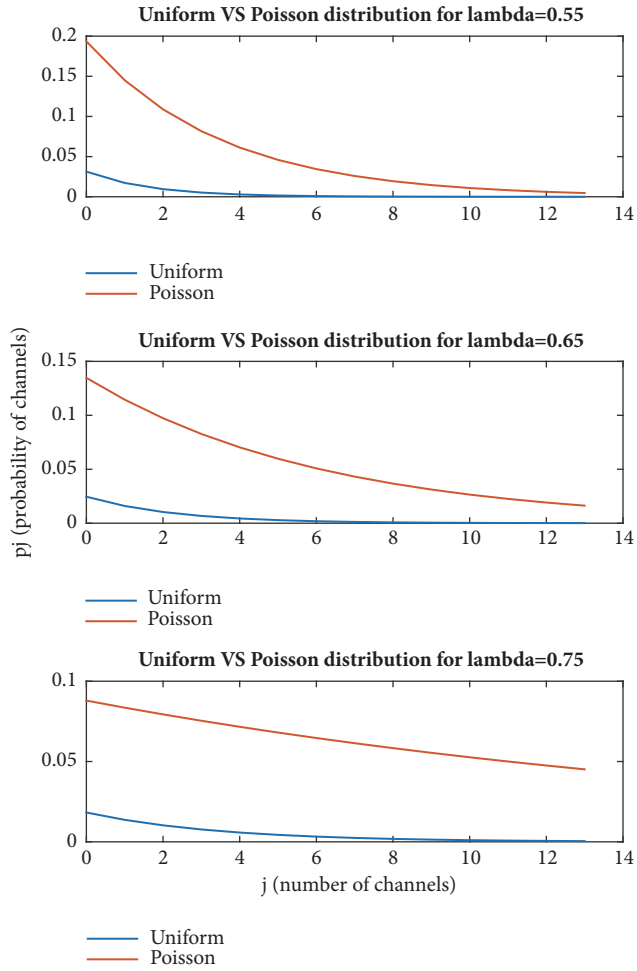


FIGURE 8: Comparison of Uniform and Poisson distributions for $\lambda = 0.55, 0.65, 0.75$.

Case 5. Case 5 presents simulation for the comparison between numerical results for Uniform and Poisson distributions using different values of λ. These results are shown in Figure 8.

Above simulation results show that an exponential-like decay in probabilities is observed in almost all the plots by increasing the number of users. This is quite in line with the expectation, i.e., considering the expression (9), and λ is kept constant which is less than 1. Hence λ^{K+1} will be negligible in comparison with 1 and

$$p_j \approx \lambda^j (1-\lambda) \sum_{k=0}^{K} \frac{P(\mathbf{k}=k)}{1-\lambda^{k+1}} \approx \lambda^j (1-\lambda) = \lambda^j \tag{13}$$

where $\sum_{k=0}^{K} P(\mathbf{k}=k) \approx 1$.

Hence by increasing j, p_j decreases exponentially as is evident from plots.

In case we take the value of λ closer to unity, the PDF becomes almost linear i.e., a straight line. It may be observed that the area under the curve which can be easily computed

in case of straight line is approximately unity that is the normalization of PDF.

6. Conclusion

Methods of finding probabilities of number of active users for existing cellular networks and for cognitive radio networks are presented. For existing cellular networks, the number of channels for base station and maximum allowable number of active users are considered to be fixed. The state of the system at any time is defined by the number of active users at that time. The probability of the system to be in a particular state is evaluated on the basis of the rate α at which a new user appears in the system and the rate β at which an active user leaves the system. Then flexible queuing state model addresses the complicated scenario for cognitive radio networks where the available channels are actually the spare channels of primary BS. Hence both the number of channels for cognitive network and the maximum number of active cognitive users are not deterministic. The probability of the cognitive network system to be in a particular state is evaluated on the basis of the rate α_c at which a new cognitive user may appear in the system, the rate β_c at which an active cognitive user may leave the system and pmf of number of available channels.

The work is important to deal with handover problem for number of active users in existing cellular networks and in cognitive radio networks on the basis of their probabilities.

Data Availability

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

Conflicts of Interest

All the authors declare that there are no conflicts of interest.

References

[1] Y.-J. Liu, S.-M. Cheng, and P.-Y. Huang, "Cognitive vertical handover in heterogeneous networks," in *Proceedings of the 11th EAI International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, QSHINE 2015*, pp. 392–397, Taiwan, August 2015.

[2] E. M. Malathy and V. Muthuswamy, "Vertical handover facilitation through queuing model in heterogeneous wireless networks," *International Journal of Advanced Engineering Technology*, vol. 7, pp. 346–348, 2016.

[3] Y. Shim and D. Shin, "Analyzing the development of 4th generation mobile network in China: actor network theory perspective," *Emerald Insight*, vol. 17, no. 1, pp. 22–38, 2015.

[4] L. A. Magagula, H. A. Chan, and O. E. Falowo, "Handover approaches for seamless mobility management in next generation wireless networks," *Wireless Communications and Mobile Computing*, vol. 12, no. 16, pp. 1414–1428, 2012.

[5] S. Alam, M. F. Sohail, S. A. Ghauri, I. M. Qureshi, and N. Aqdas, "Cognitive radio based Smart Grid Communication Network," *Renewable & Sustainable Energy Reviews*, vol. 72, pp. 535–548, 2017.

[6] S. Alam, A. N. Malik, I. M. Qureshi, S. A. Ghauri, and M. Sarfraz, "Clustering-based channel allocation scheme for neighborhood area network in a cognitive radio based smart grid communication," *IEEE Access*, vol. 6, pp. 25773–25784, 2018.

[7] H. T. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 676–688, 2011.

[8] A. Ghasemi and E. S. Sousa, "Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks," in *Proceedings of the 4th Annual IEEE Consumer Communications and Networking Conference*, pp. 1022–1026, January 2007.

[9] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Adaptive and distributed access to spectrum holes in cognitive radio system," *Wireless Personal Communications*, vol. 70, no. 1, pp. 207–226, 2013.

[10] G. I. Tsiropoulos, O. A. Dobre, M. H. Ahmed, and K. E. Baddour, "Radio resource allocation techniques for efficient spectrum access in cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 824–847, 2014.

[11] W.-C. Wu and K.-C. Chen, "Identification of active users in synchronous CDMA multiuser detection," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 9, pp. 1723–1735, 1998.

[12] A. S. Cacciapuoti, M. Caleffi, L. Paura, and M. A. Rahman, "Channel availability for mobile cognitive radio networks," *Journal of Network and Computer Applications*, vol. 47, pp. 131–136, 2015.

[13] A. L-Garcia, *Probability, Statistics and Random Processes for Electrical Engineering*, Chapter 12, Prentice Hall, 3rd edition, 2008.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2839797, 8 pages
https://doi.org/10.1155/2018/2839797

Research Article

Architecture for Collision-Free Communication Using Relaxation Technique

Saeed ur Rehman,[1,2] Rizwan Akhtar,[1] Zuhaib Ashfaq Khan,[2] and Changda Wang[1]

[1]*School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China*
[2]*COMSATS University Islamabad, Attock Campus, Pakistan*

Correspondence should be addressed to Changda Wang; changda@ujs.edu.cn

Received 1 July 2018; Revised 25 September 2018; Accepted 18 October 2018; Published 4 November 2018

Guest Editor: Sarmadullah Khan

In today's world we are surrounded by world of smart handheld devices like smart phones, tablets, netbooks, and others. These devices are based on advance technologies of multiple-input and multiple-output, Orthogonal Frequency Division Multiplexing (OFDM), and advance data reliability techniques such as forward error corrections. High data rates are among the requirements of these technologies for which turbo and low density parity check codes (LDPC) are widely used in these standards. In order to get high speed, we need multiple and parallel processors for the implementation of such codes. But there exists a collision problem as a consequence of parallel processor. This problem results in increase latency and increase of hardware complexity. In this work an approach for collision problem is presented in which network relaxation technique is used which is based on a fast clique detection. The proposed approach results in high throughput in terms of latency and complexity. Furthermore, the proposed solution is able to solve the collision problem by connecting network optimization for achieving high throughput.

1. Introduction

In modern era of an advancement in technology, different types of smart devices like smart phones, tablets, etc. are communicating with each other nodes. These devices are based on advance technologies which consist of multiple-input and multiple-output, OFDM, and advance data reliability techniques such as forward error corrections [1–9]. High data rates are among the requirements of these technologies for which turbo and low density parity check codes are widely used in these standards. In order to get high speed we need multiple/parallel processors for the implementation of such codes. Such codes are developed by using multiple processors which are linked with the data storage memory elements via connecting networks.

In turbo codes we have an algorithm which is based on iteration for the computation that improved the overall performance. So to exploit these codes with their iterative nature is a challenging task which is overcome by the use of parallel multiple processors. The required throughput is obtained by the use of multiple processors in parallel architecture. This solution increases the throughput as the latency of the system becomes the latency of the constituent subblocks [3]. Moreover, the cost of the system and the complexity are increased due to parallel nature of the architecture.

In Section 2 the state of the art is given. Section 3 gives the problem formulation. The proposed work is given in Section 4 with the pedagogical example in Section 5. Experiment and results are presented in Section 6 whereas Section 7 concludes the paper.

2. State of the Art

Different parallel processing elements depending on the interleaving law may access the same memory block simultaneously as shown in Figure 1. This architecture consists of processors linked with banks. The controller consists of control bits for connecting network and bits for address sequences of the banks. There exists memory access problem in such architectures known as the "collision" problem [7].

For solving collision problem, the existing solutions can be categorized into two families. The first family includes

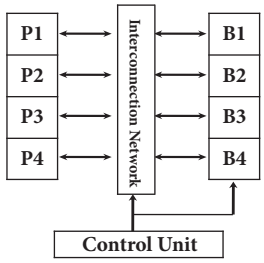


Figure 1: General architecture.

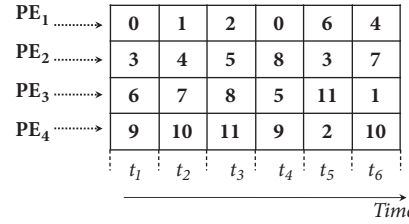


Figure 2: Memory access schedule.

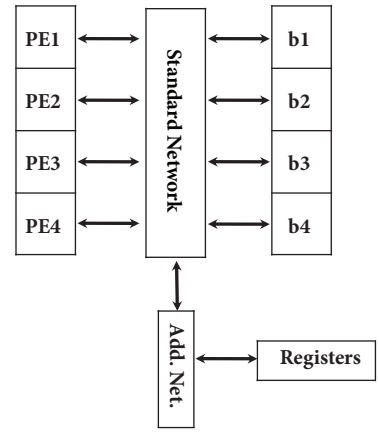


Figure 3: Architecture generated by memory adaptation.

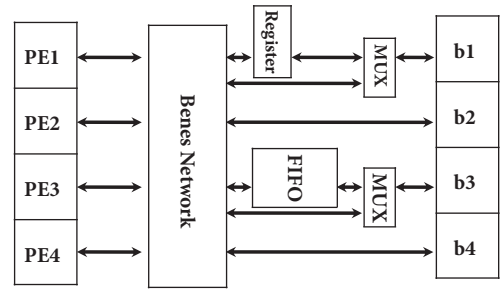


Figure 4: Architecture generated by time relaxation.

solutions like [10–14] in which the problem is solved without any changes in the standard architecture; i.e., these approaches use a standard network/memory. In [10], that is proposed with the network option but limited to certain codes. Metaheuristics approach [12] overcame this limitation but not the polynomial time. The polynomial is introduced in [13].

Works of [15–19] are of the second category which offers certain adaptation in the architecture for the storage and/or connecting network mechanism. Due to the adaptation in the real architecture the results may lack throughput in terms of latency and complexity. Reference [15] proposed to use buffers in the BENES network [20] for solving collisions. Reference [16] used advance network techniques with dedicated mechanism of routing. The resultant solution needs buffering for such mechanism as well. The work of [17] presents a network-on-chip interconnection architecture. This network allows supporting the access conflicts thanks to dedicated routing algorithm. Reference [18] is based on use of additional memory (memory adaptation) when there is no other existing solutions. But, this approach generates a huge amount of additional registers leading to expensive architectures, especially for LDPC. Since the goal of [18] is to respect a user defined network, this constraint impacts this final architectural cost on a larger scale.

Our proposal is to introduce degrees of liberty not in the memory architecture, but in the network. Hence, we will target the original source of the conflict, and we will be able to generate strongly optimized architectures. This concept will be later referred in this paper as network adaptation.

3. Problem Formulation

We will demonstrate a multiple processor parallel architecture having **P** processors which are denoted by PE which is linked with B banks of memory where number of processors are equal to number of banks. We have the size of storing data referred as D.

Figure 2 shows the example to explain the above-mentioned approach. Processors are considered as 4 which is equal to the number of banks. The time to process the data is represented in columns of the table. In first time the four processors will need 0, 3, 6, and 9 for processing as shown in Figure 2. This is a typical way to show turbo decoding. The main goal is to find collision-free mapping with specific laws and number of processors with the option of interconnection network.

The state-of-the-art architecture [12] is shown in Figure 1. The control bits will be generated in very irregular pattern having limited optimization. However, the regular pattern for such cases can be very useful for optimized results.

The regularity pattern can be seen in [18] but the finale architecture will have overhead cost due to inclusion of the logical memory elements and its connecting network (see Figure 3).

In [15], these could be avoided in which BENES network is considered in the architecture along with additional buffer as shown in Figure 4, called time relaxation resulting in latency.

In this work, an approach is presented in which a special technique is used which results in high throughput in terms of latency and complexity. The proposed solution is able

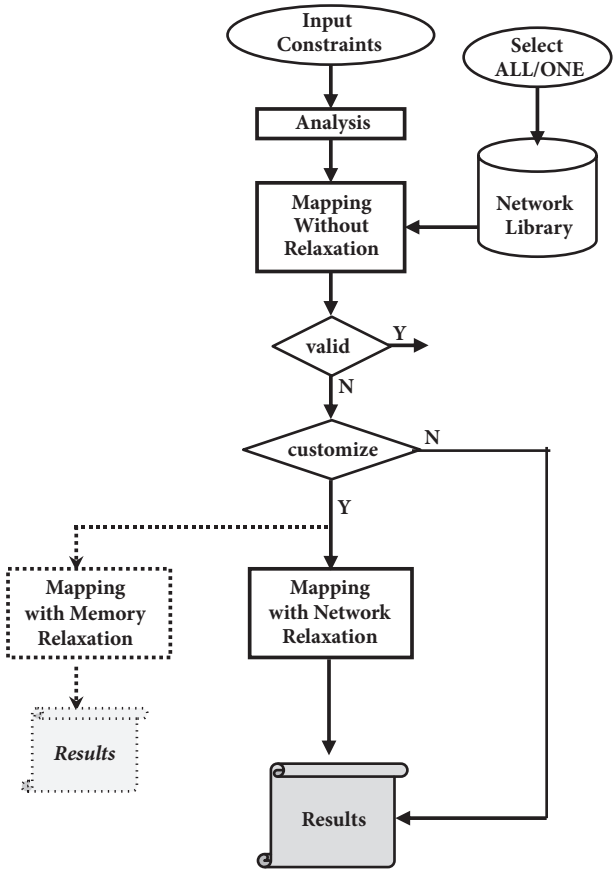


Figure 5: Proposed design flow.

to solve the collision problem by the connecting network optimization for achieving high throughput.

4. Proposed Approach

In [18], it is observed that the selected connection network has a vital role in the complexity of the proposed architecture. The proposed technique aims to use the principle based on both memory mapping space and the network, as can be seen in Figure 5. The figure presents a design space exploration for a given memory mapping problem.

The input constraint is addressed in the first step which includes laws, number of processors, and network selection. It is to determine whether a collision-free in-place memory mapping solution may exist or not for this set of constraints. This will provide useful information for the memory mapping step of our flow. The second step is to apply appropriate algorithm targeting a connecting network stored in the library. This mapping step is performed with a modified version of the mapping algorithm with memory relaxation proposed in [18]. The modifications aim to see all the collision-free mapping solution space and to stop the mapping algorithm as soon as a register should be added because of unsolvable memory mapping conflicts. If the algorithm is stopped for this specific reason, which means the selected interconnection network is too much constrained for the interleaving law.

Then, if the designer validates the customization of this interconnection network, our memory mapping algorithm with network relaxation is applied.

In the end, the cost is calculated and the results are generated. We repeat all the above process for all the available networks one by one. The user can restrict the library to explore only a subset of network(s); otherwise all the networks are explored one by one. We also integrate for experimental purpose the memory mapping approach based on memory relaxation, as given on Figure 5.

4.1. Analysis through Fast Clique Detection Algorithm (FCD). Memory mapping problem could either be based on [10] like in turbo decoders called *in-place* or be based on [18] like in LDPC decoders called *dual-memory* mapping. A fast clique detection algorithm is then proposed which determined the category of the problem.

4.1.1. Algorithm Principles. For a particular block of $\boldsymbol{L}$ data with the parallelism $\boldsymbol{P}$, our approach first constructs $\boldsymbol{L} * \boldsymbol{L}$ adjacency matrix and then fills the matrix according to the data access: if two data are accessed in the same time (e.g., data 1, 2, 3, and 4 at cycle t_1 in Figure 2), then a link between them will be stored in this adjacency matrix. This model is in fact a well-known approach used to store conflict graphs.

Next, our fast clique detection algorithm explores this matrix: if the number of interconnected data elements in a selected row is lower or equal to $\boldsymbol{P}$, then this means that data is interconnected to less than $(\boldsymbol{P}+\boldsymbol{1})$ data, so the corresponding clique, if it exists, will include less than $(\boldsymbol{P}+\boldsymbol{1})$ data, so *in-place* memory mapping remains possible; on the contrary, if this data is interconnected to at least $(\boldsymbol{P}+\boldsymbol{1})$ data (i.e., at least $(\boldsymbol{P}+\boldsymbol{1})$ of these data are involved in a common clique), this means that additional memory will be needed to keep an *in-place* memory architecture or a *dual-memory* mapping must be used. Since our goal is only to detect the existence of a clique involving at least $(\boldsymbol{P}+\boldsymbol{1})$ data, the FCD algorithm stops. In this case, we know that, in order to target architecture with no additional memory, an in-place memory mapping is not possible.

4.2. Mapping without Relaxation. Mapping without relaxation is to find a valid memory mapping for a given problem based on networks available in the library or on a user selected network. This mapping step is performed through our modified version of [18] algorithm.

The available networks in the library are barrel shifter, *BS*, butterfly, *BF*, and BENES, *BEN*. In case of exploration of all the network of the library, a memory mapping is performed targeting the lower cost network in the library (i.e., *BS* in our case), and then all other network as long as a conflict free memory mapping is possible with no relaxation. Then the lower cost architecture is proposed to the user. If not memory mapping is possible, then no solution can be proposed to the user at this step. It must be noticed that *BS* or *BF* does not guarantee a valid memory mapping for a given problem, but a fully connected *BEN* network can always give a valid memory mapping.

```
Set_perm    // permutations that exists in the network
Map // The matrix of mapping
Col_i;      // The Map column which is selected
Pcol_i;  // Permutation Col_i
Dpcol_i; // One data of Pcol_i
Algorithm NetworkMap (set_Perm, Map, Startup)
Coli = SelectColumn(Map);
PColi = SelectValidPerm(set_Perm, Coli, Map);
If ((PColi is not empty) and not (ALLMAP(MAP))) Then
    Do
      DPColi = Select&RemoveFirstPerm(PColi);
      MapColumn(Coli, DPColi);
      Startup = FALSE;
      NetworkMap (set_Perm, Map, Startup);
      If ((Startup = FALSE) and not(ALLMAP(MAP))) Then
         RemoveColumn(Coli, DPColi);
      End if;
    While ((Startup = FALSE) and (PColi is not empty) and
         not(ALLMAP(MAP)));
    If ((Startup = FALSE) and not(ALLMAP(MAP))) Then
        AddNewNetworkComp(set_Perm);
        EraseMap(Map);
        Startup = TRUE;
        NetworkMap (set_Perm, Map, Startup);
    End if;
End if;
```

Pseudocode 1

4.3. Mapping with Network Relaxation. Mapping with network relaxation gives the liberty to change the network architecture to find a valid memory mapping by adding network components. If mapping is not possible with the selected standard network, certain additional network will be added to the standard network which will result in a valid mapping. The mapping adaptation is given by adaptation the network selected in the beginning with the addition in terms of some elements such as switches and/or multiplexers. This aim to take advantage of [10, 12] for high throughput cost and latency, in such a way that the application used in [15, 18] can be targeted.

For a standard network, the mapping algorithm finds a memory mapping according to the permutation available for the network in the library. If the mapping is not possible for the targeted network an additional network component (e.g., switch) is added to the network. The new interconnection network architecture composed of standard network with additional switch generates new set of permutations. The mapping algorithm is applied based on the new set of permutations. This approach keeps on adding new network component until a collision-free mapping is obtained.

Figure 6 shows the architecture with network relaxation which is composed of original connecting network having adaptation of additional components. As memory adaptation is achieved so optimized architecture can be obtained as compared to [15] or [18], since memory elements are more costly than MUX/switches.

The pseudocode of the proposed approach is given as shown in Pseudocode 1.

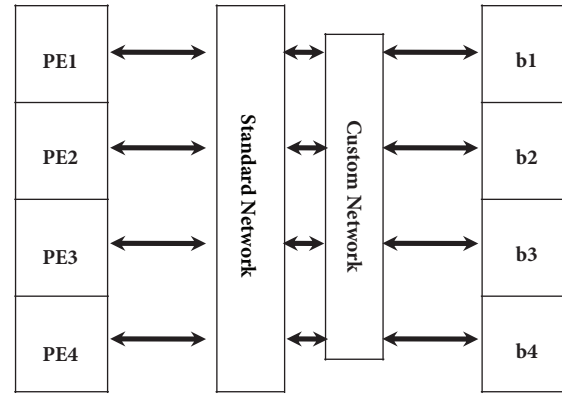


Figure 6: Architecture generated by network relaxation.

5. Pedagogical Example

The proposed approach presented in Section 3 is explained with a small example of size **L**=12 and parallelism **P**=4 from high speed packet access (HSPA) turbo decoders architecture is considered.

The whole design space is explored using all the networks one by one available in the library. The data access pattern is shown in Figure 7.

5.1. Analysis. First of all the clique test is carried out in order to find out whether the collision problem is based on in-place or double memory mapping. As $\boldsymbol{L} = \mathbf{12}$ and $\boldsymbol{P} = \mathbf{4}$, construct

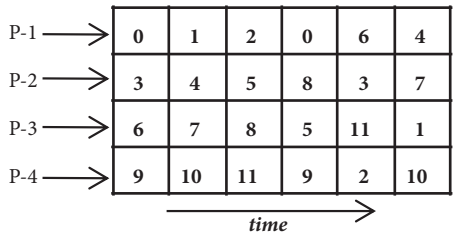


Figure 7: Pedagogical example.

	0	1	2	3	4	5	6	7	8	9	10	11
0	x			x		x	x		x	x		
1		x			x			x			x	
2			x	x		x	x		x			x
3	x		x	x			x			x		x
4		x			x			x			x	
5	x		x			x			x	x		x
6	x		x	x			x			x		x
7		x			x			x			x	
8	x		x			x			x	x		x
9	x			x		x	x		x	x		
10		x			x			x			x	
11			x	x		x	x		x			x

Figure 8: Clique test matrix.

12*x*12 matrix and fill the matrix according to the data access. In Figure 8, the complete matrix is shown filled according to all the access of Figure 7.

Now starting from the first row as it can be seen that 6 data elements are connected with each other which are greater than ***P***, so we will find the number of combinations and determine the clique. After checking all the rows no such clique is found so the problem is based on in-place memory mapping.

The considered example is shown in Figure 9 such that data can be mapped according to the described mapping algorithm.

5.2. Mapping without Relaxation. The mapping algorithm [10] is applied based on all the networks available in the library one by one.

Firstly, the data matrix is mapped based on BS. For considered example collision-free solution is not possible as none of the *BS* permutations are valid for mapping as shown in Figure 10(a). The next available network in the library to find the mapping is *BF*. The whole matrix is mapped successfully for BF network. The resultant mapping is shown in Figure 10(b). Finally, memory is mapped with *BEN*.

The given problem cannot be mapped using *BS* so network and memory relaxation methods can be considered to find collision-free mapping with targeted network.

5.3. Memory Relaxation. The memory relaxation approach adds registers in the architectures to solve the data with collision and memory locations when there is no other solution. Figure 11 shows a complete mapping consist of registers.

5.4. Network Relaxation. Network relaxation gives the liberty to modify the network architecture. As shown in Figure 10(a), for the last column none of the permutation from BS is supported. So an additional network component is added to the standard BS network and new permutations are generated. Continue the mapping process by using the current new set of permutations. The resultant mapping is shown in Figure 12.

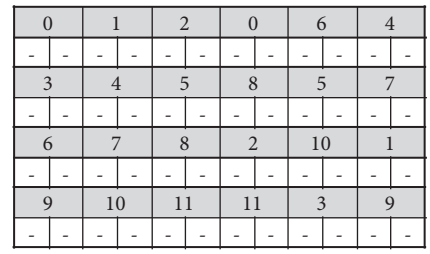
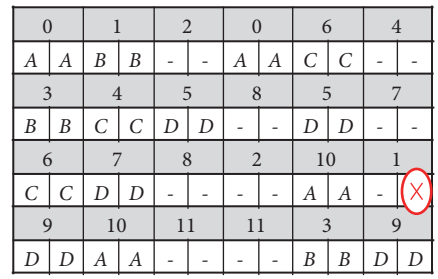
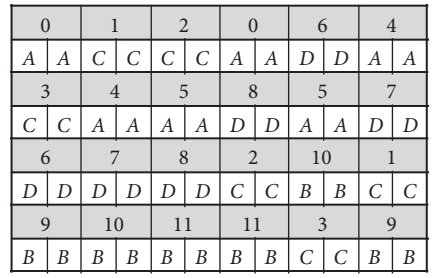


Figure 9: Matrix representation.



(a) Targeting barrel shifter



(b) Targeting BF

Figure 10: Mapping without relaxation.

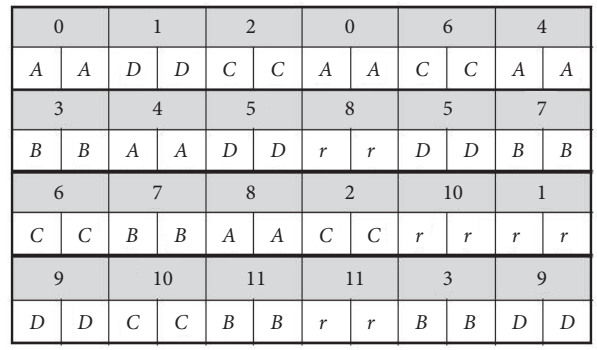


Figure 11: Mapping with memory relaxation.

Table 1: Area and latency comparison.

	Area [15]	Area [12]	Area Proposed NW relax.	Latency [15]	Latency [12]	Latency Proposed NW relax.
L=40	21386	31666	15849	44	20	20
L=120	78786	129246	64639	134	60	60
L=800	690096	1090696	545364	1000	400	400
L=1560	1456386	2574486	1343224	1560	780	780
L=2240	2090656	3696656	1848344	2800	1120	1120

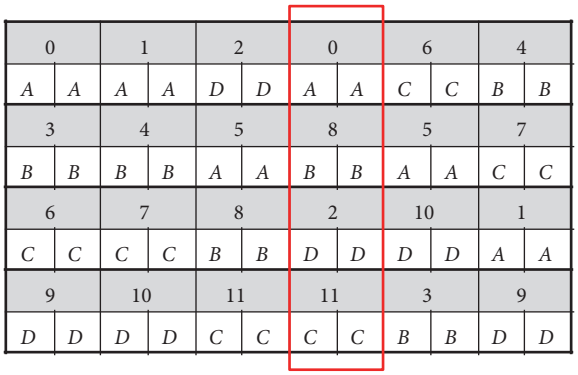
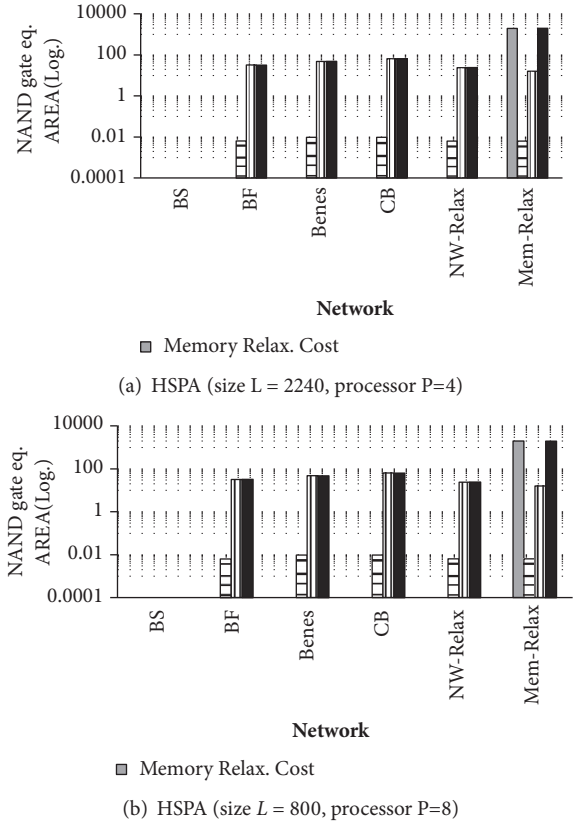


Figure 12: Mapping with network relaxation.



(a) HSPA (size L = 2240, processor P=4)

(b) HSPA (size L = 800, processor P=8)

Figure 13: Cost comparison for different networks.

6. Experiments and Results

The experiments and results performed are presented in this section. STMicroelectronics 90 nanometer technology is used in these experiments to determine the NAND-gate equivalent area of various of the architecture.

This approach can be used for any type interleaver with any number of selected processors for parallelism. We have considered implementation of interleavers that are in use of HSPA Evolution [1].

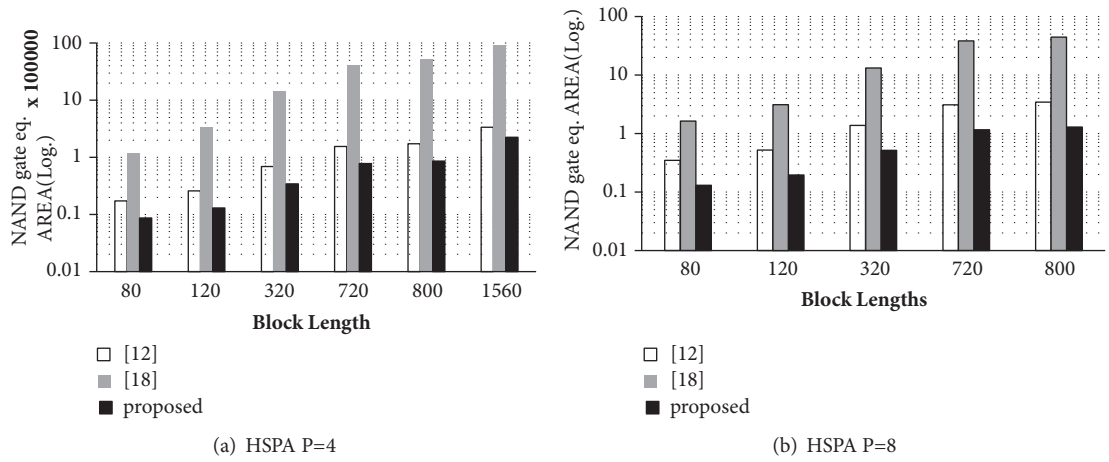
Table 1 shows assessment of the network adaptation approach with different other approaches presented in the literature. The results are shown as comparison of NAND-gate equivalent area and latency in terms of cycles for many HSPA block sizes $\boldsymbol{L}$ with number of processors $\boldsymbol{P}$ = 4.

The network adaptation approach has above half of lesser area having similar latency as compared to [12] and 20% area is lesser with reduction of 55% in latency as compared to [15]. Hence, the network adaptation approach can be used to reduce the cost significantly as compared to [12] and it reduces the latency compared to [15].

In order to explore the proposed approach an example of HSPA with Block length of 2240 is considered. Figure 13 shows results for the given problem based on BS, BF, and BENES network excluding the memory cost which is considered same for all the results. Figure 13 also includes results with memory and network relaxation methods.

As the given problem cannot be mapped using *BS*, the cost for memory relaxation using a *BS* is very high as compared to newly proposed network relaxation method. Moreover, we have taken different block sizes using HSPA standard with number of processors $\mathbf{P}$ = 4 & $\mathbf{P}$ = 8. Figure 14(a) shows the results for four processors and Figure 14(b) shows 8 processors. The result clearly shows that the network adaptation method gives optimal results. Network relaxation for a targeted network can reduce the cost of controller 400 times as compared to existing memory adaptation approach.

In Figure 15, the comparison of the proposed approach is presented with [12, 18, 19]. The results showed significant reduction is cost as compared to these approaches. In [21], the authors proposed a solution for the standard LTE which is based on multistage connecting network of barrel shifter. For 4 processors, two stages of barrel shifter with modification need 3 bits while for $\mathbf{P}$ = 8, three stages of modified *BS* need 7 bits. Experiments of the proposed generalized approach are able to find some block lengths that support standard *BS*.



(a) HSPA P=4 (b) HSPA P=8

Figure 14: Cost comparison for different block lengths.

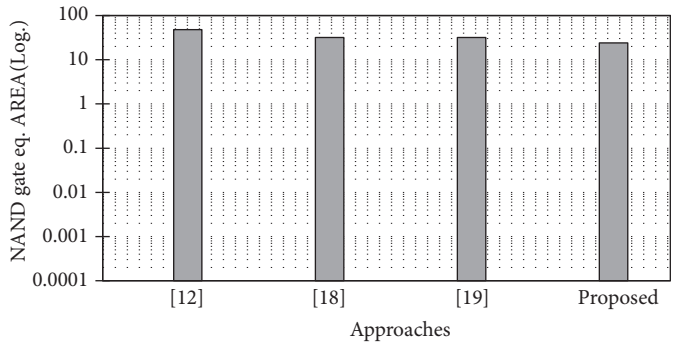


Figure 15: Cost comparison for different block lengths.

Table 2: Block lengths supported by BS.

P=4	160,200,240,320,360,480,528,800,880, 960,1440,1600,2240,2880,3520,4160, 4480,5760,4160
P=8	416,480,800,1600,2240,3520,3584, 4608,4480

The standard barrel shifter uses 2 bits for $\mathbf{P} = 4$ and only 3 bits for $\mathbf{P} = 8$. This will reduce the network controller cost up to 133%. Table 2 shows *LTE* block lengths that support *barrel shifter*.

7. Conclusion

In this work we have presented an approach in which a special technique is used which results in high throughput in terms of latency and complexity. The proposed solution is able to solve the collision problem by the connecting network optimization for achieving high throughput. The result clearly showed that the network adaptation method gives optimal results. Network adaptation for a targeted network can reduce the cost of controller 400 times as compared to existing memory adaptation approach. It can be noted that the network controller cost mostly affects the total cost of the architecture; therefore, in future the controller cost could be reduced to get more optimized results. Furthermore, other targeted networks shall be studied in order to get optimized results.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Science Foundation of China under Grant 61672269 and the Jiangsu Provincial Science and Technology Project under Grant BA2015161.

References

[1] Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access; Multiplexing and Channel Coding (Release 8), *3GPP Std. TS 36.212*, Dec. 2008.

[2] 3GPP, "Technical specification group radio access network multiplexing and channel coding (FDD)" (25.212 V5.9.0). June 2004.

[3] *DVB DocumentA122*. "Frame structure channel coding and modulation for the second generation digital terrestrial television broadcasting system (DVB-T2)," 2008.

[4] A. Giulietti, L. Van Der Perre, and M. Strum, "Parallel turbo coding interleavers: Avoiding collisions in accesses to storage elements," *IEEE Electronics Letters*, vol. 38, no. 5, pp. 232–234, 2002.

[5] *IEEE P802.11n/D5.02, Part 11*. "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Enhancements for Higher Throughput", July 2008.

[6] *IEEE P802.16e, Part 16*. "Air Interface for Fixed and Mobile Broadband Wireless Access Systems," Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, and Corrigendum 1, Feb. 2006.

[7] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *IEEE Electronics Letters*, vol. 32, no. 18, pp. 1645-1646, 1996.

[8] A. I-Gutierez, J.-B. Mourade, S.-P. fletschinger et al., "DAVINCI Non-Binary LDPC codes: Performance and Complexity Assessment," in *Proceedings of the future Network & Mobile Summit, Italy*, June 2010.

[9] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near-Shannon limit error-correcting coding and decoding: Turbo codes," in *Proceedings of the IEEE International Conference on Communications*, vol. 2, 1993.

[10] C. Chavet, C. Phillippe, M. Eric et al., "Static Address Generation Easing: a Design Methodology for Parallel Interleaver Architectures," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1594–1597, Dallas, March 2010.

[11] L. Dinoi and S. Benedetto, "Variable-Size Interleaver Design for Parallel Turbo Decoder Architectures," *IEEE Transactions on Communications*, vol. 53, no. 11, pp. 1833–1840, 2005.

[12] A. Tarable, S. Benedetto, and G. Montorsi, "Mapping interleaving laws to parallel turbo Low density parity check decoder arch," *IEEE Transactions on Information Theory*, 2004.

[13] A. H. Sani, P. Coussy, and C. Chavet, "A first step toward on-chip memory mapping for parallel turbo and LDPC decoders: a polynomial time mapping algorithm," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4127–4140, 2013.

[14] T.-C. Wu, C.-M. Lee, and C.-K. Wang, "Construction of parallelized-decoding LDPC codes," in *Proceedings of the 2016 IEEE International Symposium on Circuits and Systems, ISCAS 2016*, pp. 425–428, Canada, May 2016.

[15] N. When, *SOC-Network for Interleaving in wireless Communications*, MPSOC, 2004.

[16] O. Muller, A. Baghdadi, and M. Jézéquel, "ASIP-based multiprocessor SoC design for simple and double binary turbo decoding," in *Proceedings of the Design, Automation and Test in Europe, DATE'06*, Germany, March 2006.

[17] H. Moussa, A. Baghdadi, and M. Jézéquel, "Binary de Bruijn on-chip network for a flexible multiprocessor LDPC decoder," in *Proceedings of the 45th DAC annual conference*, p. 429, Anaheim, California, June 2008.

[18] Ar. Briki, Ch. Cychile, and C. Philipe, "A Memory Mapping Approach for Network and Controller Optimization in Parallel Interleaver Architectures," in *Proceedings of the 2013 IEEE Workshop on Signal Processing Systems (SiPS)*, 2013.

[19] E. Nieminen, "On quadratic permutation polynomials, turbo codes, and butterfly networks," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 63, no. 9, pp. 5793–5801, 2017.

[20] V. E. Benes, *Mathematical theory of connecting networks and telephone traffic*, Academic Press, New York, NY, USA, 1965.

[21] C.-C. Wong and H.-C. Chang, "Reconfigurable turbo decoder with parallel architecture for 3GPP LTE system," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 7, pp. 566–570, 2010.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3153915, 14 pages
https://doi.org/10.1155/2018/3153915

Research Article

One-to-Many Relationship Based Kullback Leibler Divergence against Malicious Users in Cooperative Spectrum Sensing

Noor Gul,[1,2] **Ijaz Mansoor Qureshi**,[3] **Sadiq Akbar**,[2] **Muhammad Kamran**,[2] **and Imtiaz Rasool**[2]

[1]*Department of Electrical Engineering, Faculty of Engineering and Technology, International Islamic University, Islamabad 44000, Pakistan*
[2]*Department of Electronics, University of Peshawar, Peshawar 25000, Pakistan*
[3]*Department of Electrical Engineering, Air University, Islamabad 44000, Pakistan*

Correspondence should be addressed to Noor Gul; noor.phdee51@iiu.edu.pk

Received 21 April 2018; Revised 2 August 2018; Accepted 11 August 2018; Published 2 September 2018

Academic Editor: Sungchang Lee

The centralized cooperative spectrum sensing (CSS) allows unlicensed users to share their local sensing observations with the fusion center (FC) for sensing the licensed user spectrum. Although collaboration leads to better sensing, malicious user (MU) participation in CSS results in performance degradation. The proposed technique is based on Kullback Leibler Divergence (KLD) algorithm for mitigating the MUs attack in CSS. The secondary users (SUs) inform FC about the primary user (PU) spectrum availability by sending received energy statistics. Unlike the previous KLD algorithm where the individual SU sensing information is utilized for measuring the KLD, in this work MUs are identified and separated based on the individual SU decision and the average sensing statistics received from all other users. The proposed KLD assigns lower weights to the sensing information of MUs, while the normal SUs information receives higher weights. The proposed method has been tested in the presence of always yes, always no, opposite, and random opposite MUs. Simulations confirm that the proposed KLD scheme has surpassed the existing soft combination schemes in estimating the PU activity.

1. Introduction

The rapid evolution in wireless communication demands new wireless services in both the used and unused parts of the radio spectrum [1]. The Federal Communications Commission (FCC) exclusively assigns spectrum bands to various services [2, 3]. Cognitive radio (CR) is a smart technique that gains knowledge from the environment and adjusts its parameters accordingly [4]. The incumbent primary users (PUs) are free to transmit any time with no restrictions, while the secondary users (SUs) can get the benefit of the spectrum only when the licensee declares it free [5].

In cognitive radio network (CRN), sensing the incumbent user spectrum is vital. An offensive detection on the PU channel due to false alarm reduces the SUs opportunity to utilize the free spectrum. Similarly, any misdetection in the PU transmission will produce interference in the transmission of legitimate and opportunistic users. In case of the frequent usage of the spectrum by the PUs, the termination probability of SUs is not easy to ensure. The proposed scheme in [6] uses channel reservation to improve the quality of service (QoS) for SUs. To confirm the status of the PU a number of detection schemes, such as energy detector (ED), feature detector (FD), and matched filter detector (MFD), are used in [3, 7]. The use of multiple antennas for spectrum sensing is considered in [8], when noise and signal of the PU are considered as independent complex zero-mean Gaussian random variables. Detection results of the orthogonal frequency division multiplexing (OFDM) signals in the frequency selective fading channels are considered in [9, 10]. Optimal interference subcarriers are obtained based on Genetic Algorithm (GA) to suppress intercarrier interference of the unlicensed user to the licensee [11]. In [12], side-lobes reduction using generalized side-lobe canceller combined with GA and differential evolution is proposed. The GA and interior point method design schemes enhance

the performance of hybrid computing technique based on active noise control for random, complex random, and sinusoidal input noise variations in the primary secondary paths [13]. An improved energy detector scheme is suggested in [14] to maximize the throughput of a CRN network with minimum interference to the PUs.

Individual SU faces a number of restrictions to sense the PU spectrum accurately. The sensing performance of a solitary device is degraded due to fading, shadowing, energy constraints, and hidden station dilemmas of the primary signal. It is therefore most probable that a single user might fail to spot detection of the licensee transmitter [5]. The cooperative user devices placed more than a few wavelengths apart experience an independent fading effect. The doubt to efficiently detect the licensed spectrum possession is mitigated by enabling different users to share their local sensing results and make a cooperative decision [15–19]. The particle swarm optimization is able to find their foods by sharing their search information in contrast with GA using crossover and kids reproduction [20]. Cooperative spectrum sensing (CSS) using energy harvest-based weighed method in [21] reduces the energy wastage of the SUs with better sensing results. The use of cyclostationary signatures in [22] is used to subdue large number of challenges related to cooperative network in the rising applications of the CR. The proposed method in [23] enhances throughput of the CRN in the presence of mobile SUs to access the PU spectrum. The MUs presence in a CSS environment reduces the effectiveness of the cooperation. Therefore, precise recognition and exclusion of the false sensing information are extremely vital [24]. Significant investigations are carried out to make the collaborative schemes resistant to any MU attack. The aim of any MU in CSS is to provide false sensing data to the FC [25].

The assistance of trusted nodes in a reputation based CSS network is discussed in [26], where Nyman-Pearson and likelihood ratio test is utilized for spectrum sensing improvements. The primary emulation category of MUs discussed in [27] tries to impersonate activities of the legitimate PU transmitter. A robust technique with prime focus on always yes group of MUs is implemented in [28]. An extended sequential cooperative scheme that reduces the number of sensing reports is investigated in [29]. In the soft fusion combination schemes proposed in [30–32] sensing energies from different SUs are combined to take accurate decision about the PU spectrum holes. Similarly, in the hard fusion schemes SUs provide a hard binary decision to the FC to predict the licensed user activity in the spectrum [33–35]. The optimal quantization scheme in [36, 37] is able to produce improved detection with a control on the probability of false alarm. Bioinspired heuristics based on GA find the design parameters of nonlinear Hammerstein controlled autoregressive systems with distinct values of the noise variance [38]. The study in [39–45] focuses on evolutionary computation for optimizing the detection and false alarm probabilities to minimize the sensing error for a particular SU.

In this paper, Kullback Leibler Divergence (KLD) has been employed to protect the CSS against the spectrum falsification attack (SFA) of always no (AN), always yes (AY), random opposite (RO), and the always opposite (AO) categories of MUs by assigning weights to the sensing reports of SUs before global combination at the FC. In our previous study [46], SUs perform their local sensing, report soft energies to the FC, and also store this information in its local database. FC determines the KL divergence score against each user and also acknowledges this same information to the user. A normally declared user tries to send mean of the previous energy reports to the FC based on its current observation. The work in [46, 47] uses the KLD to determine the probability distribution function (PDF) dissimilarity of a particular SU under the presence and absence hypothesis of the licensed user channel. The PDF uses the energy statistics of an individual user under both hypotheses for declaring it normal or malicious. In this work, FC takes sensing data from all SUs and determines KLD score based on the energy statistics of individual user with the average statistics received from all other users. The final decision is made at FC by assigning weights to the local energy information of each individual SU based on the measured KLD score of each SU. Lower weights are assigned to the sensing data of MUs based on the KLD results, while the regular user sensing information receives higher weights. The lower weights keep the FC final decision less prone to the attack of MUs. By following the proposed method, the performance of CSS is kept at its maximum in the presence of MUs without identifying any malicious activity.

The proposed method results are tested in the company of AO, RO, AY, and AN categories of MUs in a cooperative environment. The outcome shows that these MUs in CSS increase the false alarm and misdetection, resulting in an increased interference to the primary transmission and reduced throughput of the network. Simulations confirmed that the proposed one-to-many relation based KLD method leads to more accurate and sophisticated detection than the traditional soft combination schemes in [46, 47].

The rest of the paper is organized as follows: In Section 2, the system model is presented. Section 3 explains the proposed scheme, where proposed method is used to overwhelm the MUs effects in the global decision of the FC. Experimental outcomes are presented in Section 4. Section 5 concludes the paper.

2. System Model

In the CSS as in Figure 1, all SUs report their local sensing information of the PU channel to the FC. The FC collects sensing notifications of all individual SUs and generates a global decision to show the actual status of the PU spectrum.

The spectrum sensing decisions H_1 and H_0 made by each SU in a particular spectrum are as follows:

$$y_j(l) = \begin{cases} H_0 & n_j(l) \\ H_1 & h_j s(l) + n_j(l) \end{cases} \quad (1)$$

where H_0 is the hypothesis about the availability and H_1 is the hypothesis for the occupancy of the PU spectrum by the licensed user. $y_j(l)$ is the received signal by the j^{th} user at the l^{th} time slot. $n_j(l)$ is the Additive White Gaussian Noise at the j^{th} user receiver. h_j is the amplitude of the channel gain and $s(l)$ denotes the PU transmit signal.

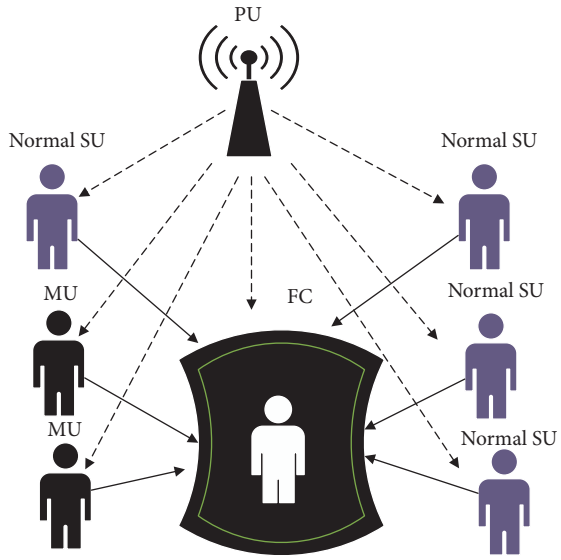


Figure 1: The conventional CSS mechanism.

Accordingly, as a consequence of H_1 and H_0 hypothesis, the observed signal energy at the j^{th} receiver can be represented as

$$E_j(i) = \begin{cases} \sum_{l=l_i}^{l_i+K-1} \left|n_j(l)\right|^2, & H_0 \\ \sum_{l=l_i}^{l_i+K-1} \left|h_j s(l) + n_j(l)\right|^2, & H_1 \end{cases} \quad (2)$$

In (2), each of the i^{th} sensing intervals is divided into K total number of samples. When the number of samples is considered large enough, the soft energy report of these SUs is similar to that of a Gaussian random variable in the H_0 and H_1 hypothesis as in [29, 30].

$$E_j \sim \begin{cases} N\left(\mu_0 = M, \sigma_0^2 = 2M\right), & H_0 \\ N\left(\mu_1 = M\left(\eta_j + 1\right), \sigma_1^2 = 2M\left(\eta_j + 1\right)\right), & H_1 \end{cases} \quad (3)$$

Here η_j is the signal-to-noise ratio between the primary transmitter and the j^{th} user. Moreover, (μ_0, σ_0^2) and (μ_1, σ_1^2) are the energy distributions means and variances when H_0 and H_1 hypotheses are true.

The KLD between the two normally distributed functions $a(x)$ and $b(x)$ is calculated as follows [46]:

$$K(a \parallel b) = \int a(x) \log\left(\frac{a(x)}{b(x)}\right) dx \quad (4)$$

Similarly, the KLD representation for functions $a(x)$ and $b(x)$ with the means and variances (μ_a, σ_a^2) and (μ_b, σ_b^2) is as follows:

$$K(a \parallel b) = K\left(\mu_a, \mu_b, \sigma_a^2, \sigma_b^2\right) = \frac{1}{2}\left(\log\left(\frac{\sigma_b^2}{\sigma_a^2}\right) - 1 + \left(\frac{\sigma_a^2}{\sigma_b^2}\right) + \frac{\left(\mu_a - \mu_b\right)^2}{\sigma_b^2}\right) \quad (5)$$

The MUs in Figure 2 are producing dissimilar energy distribution in the H_1 and H_0 hypothesis as compared with normal SUs. The KLD score against these MUs is dissimilar to the normal SUs and is easily separable from the normal user category.

The probability distribution functions of the energy statistics reported by the normal SU, AY, AN, AO, and RO users are given in Figure 2. The energy distributions provided by all four categories of MUs are different from the normal user distributions. Therefore, any cooperative user having energy distribution dissimilar to the normal user in Figure 2(a) is treated as malicious one. The AO user distribution in Figure 2(b) always negates the distribution of the normal user. The AY user in Figure 2(c) is producing similar high energy distributions under both hypotheses. Similarly, the AN user with always free state information of the licensee channel has its low energy distributions in Figure 2(d). The RO user behaves as AO with probability P and as a normal user with probability (1-P) in Figure 2(e).

3. The Proposed One-to-Many Relations Based KLD

The proposed work considers total cooperative users larger in number compared with MUs. All the cooperative users inform FC about their local spectrum observations of the primary channel. FC collects and takes its global decision based on the received energy statistics of the reporting users. Before making any global decision, FC assigns weights to the local sensing of SU reports with the proposed KLD method. The resultant weights illustrate reliability of the local spectrum sensing information of the individual cooperating users prior to making any final decision at the FC.

A pseudocode showing the proposed KLD algorithm for the local detection determining KLD score using one-to-many relationship based energy statistics and taking global decision based on the received energy and measured weights is given below:

(1) For $i = 1$ to limit

(2) For $j = 1$ to SU

(3) Local detection $E_j(i)$ by the j^{th} user

(4) New values of mean and variance $(\mu_{ja}(i), \sigma_{jb}^2(i))$ based on $E_j(i)$

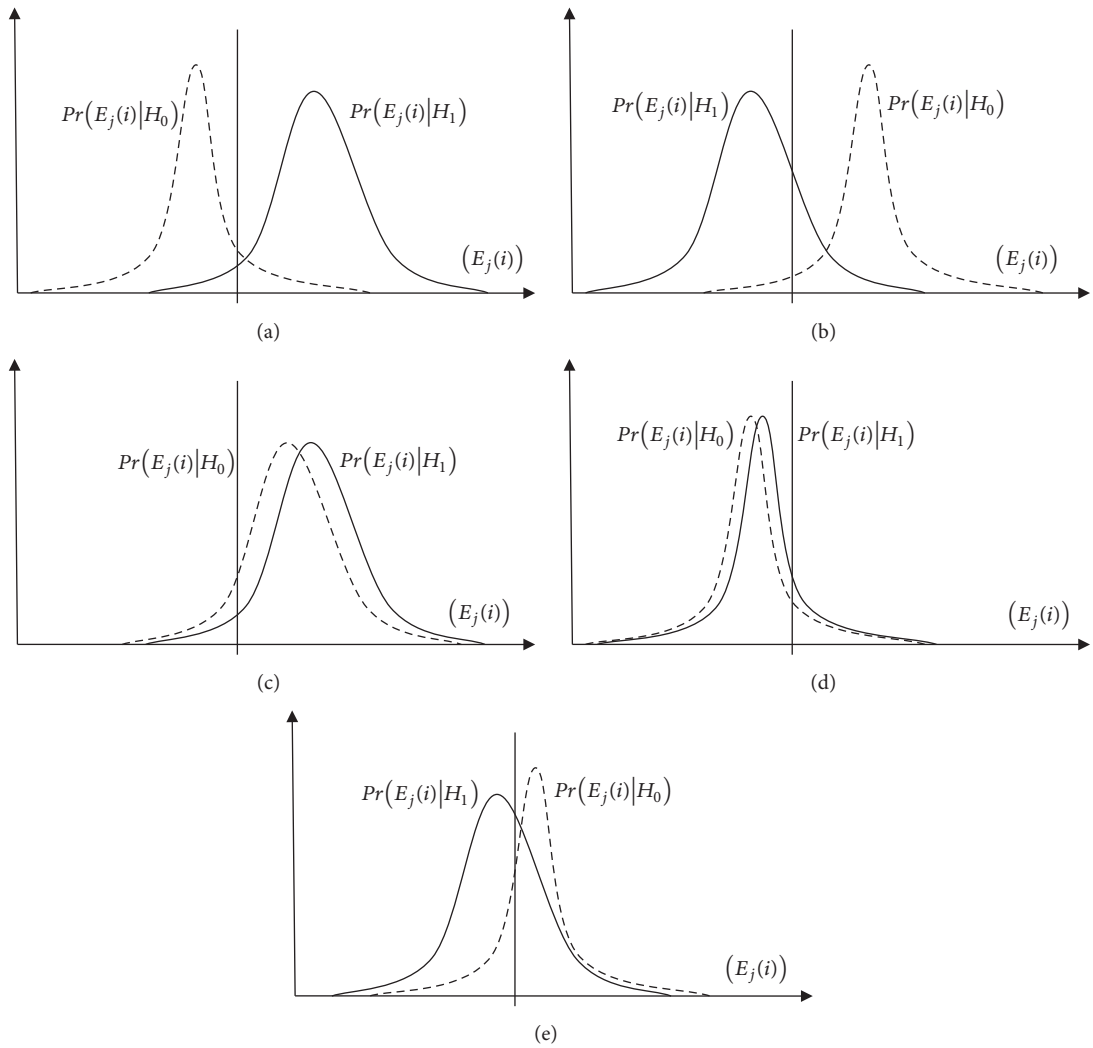


Figure 2: The probability density function of the users based on the reported statistics: (a) normal, (b) AO malicious, (c) AY malicious, (d) AN malicious, and (e) RO malicious.

(5) Average means and variance values while taking out the j^{th} user energy statistics.

$$\mu_{ja'}(i) = \left(\frac{\left(\sum_{j=1}^{M} \mu_{ja}(i)\right) - \mu_{ja}(i)}{(M-1)} \right), \quad i \in 1, \ldots N$$
$$\sigma^2_{jb'}(i) = \left(\frac{\left(\sum_{j=1}^{M} \sigma^2_{jb}(i)\right) - \sigma^2_{jb}(i)}{(M-1)} \right), \quad i \in 1, \ldots, N \tag{6}$$

(6) One-to-many relationship based KLD

$$K_j(i) = KL\left(\mu_{ja'}(i), \mu_{ja}(i), \sigma^2_{jb'}, \sigma^2_{jb}\right) \tag{7}$$

(7) Weights for the j^{th} user in the i^{th} interval

$$\boldsymbol{c}_j(i) = \left(\frac{1}{K_j(i)} \right)$$
$$\boldsymbol{w}_j(i) = \left(\frac{\boldsymbol{c}_j(i)}{\sum_{j=1}^{M} \boldsymbol{c}_j(i)} \right), \quad i \in 1, \ldots, N \tag{8}$$

(8) End SUs

(9) If $\sum_{j=1}^{M} \boldsymbol{w}_j(i) * E_j(i) \geq \varepsilon$

(10) Global decision, $G_B(i) = H_1$

(11) Else

(12) Global decision, $G_B(i) = H_0$

(13) End

(14) End limit

3.1. Data Collection and Adjustments by the FC. FC receives the individual soft energy information $E_j(i)$ in the i^{th} interval from all the j^{th} cooperating SUs as

$$\boldsymbol{e} = \left[E_1(i)\ \ E_2(i)\ \ E_3(i)\ \ \dots\ \ E_M(i)\right], \quad i \in 1, \dots, N \tag{9}$$

where $\boldsymbol{e}$ is a row vector containing the soft spectrum sensing data of all M users during the i^{th} interval. The soft energy report $E_j(i)$ has mean and variance (μ_1, σ_1^2) under hypothesis H_1 and (μ_0, σ_0^2) under the H_0 hypothesis.

FC further determines new values of the mean and variance for all users in the i^{th} sensing interval based on the received energy observations in (9) as

$$\boldsymbol{a}(i) = \left[\mu_{1a}(i)\ \ \mu_{2a}(i)\ \ \mu_{3a}(i)\ \ \dots\ \ \mu_{Ma}(i)\right], \quad i \in 1, \dots, N \tag{10}$$

$$\text{where } \mu_{ja}(i) = \begin{Bmatrix} z_1\mu_{j1} + z_2E_j(i), & H_1 \\ z_1\mu_{j0} + z_2E_j(i), & H_0 \end{Bmatrix} \tag{11}$$

Here $\mu_{ja}(i)$ is the new value of the mean for the j^{th} SU in the i^{th} sensing interval, which is updated according to the received energy $E_j(i)$ and (z_1, z_2) preselected constants.

Similarly, new variance values are determined and collected based on the received energy $E_j(i)$ as

$$b(i) = \left[\sigma_{1b}^2\ \ \sigma_{2b}^2\ \ \sigma_{3b}^2\ \ \dots\ \ \sigma_{Mb}^2\right], \quad i \in 1, \dots, N \tag{12}$$

$$\text{where } \sigma_{jb}^2(i) = \begin{Bmatrix} z_1\sigma_{j1}^2 + z_1\left[E_j(i) - \mu_{j1}\right]^2, & H_1 \\ z_1\sigma_{j0}^2 + z_1\left[E_j(i) - \mu_{j0}\right]^2, & H_0 \end{Bmatrix} \tag{13}$$

In the new mean and variance measurements in (11) and (13), the constant $z_1 = (k-1)/(k)$ and the constant $z_2 = (1/k)$, where the constant k is the effective level of the mean and variance by the received energy $E_j(i)$.

3.2. One-to-Many Relationship Based KLD Measurement. After the collection of mean and variance information on behalf of all M users in the i^{th} sensing intervals, FC measures a difference in the mean and variance of the j^{th} user energy statistics with all other users. The average mean values are measured on behalf of all M SUs based on the new mean values of (10) as

$$\mu_{ja'}(i) = \left(\frac{\left(\sum_{j=1}^{M}\mu_{ja}(i)\right) - \mu_{ja}(i)}{(M-1)}\right) \tag{14}$$

The one-to-many difference results of the mean for all M SUs are collected as

$$\boldsymbol{a}'(i) = \left[\mu_{1a'}(i)\ \ \mu_{2a'}\ \ \dots\ \ \mu_{Ma'}(i)\right], \quad i \in 1, \dots, N \tag{15}$$

Similarly, the average variance values are measured on behalf of all M SUs based on the new variance values of (12) as follows:

$$\sigma_{jb'}^2(i) = \left(\frac{\left(\sum_{j=1}^{M}\sigma_{jb}^2(i)\right) - \sigma_{jb}^2(i)}{(M-1)}\right) \tag{16}$$

$$\boldsymbol{b}'(i) = \left[\sigma_{1b'}^2(i)\ \ \sigma_{2b'}^2(i)\ \ \sigma_{3b'}^2(i)\ \ \dots\ \ \sigma_{Mb'}^2(i)\right], \quad i \in 1, \dots, N \tag{17}$$

Here $\mu_{ja'}(i)$ is the average mean and $\sigma_{ja'}^2(i)$ is the average variance value of the energy samples provided by all other users while ignoring the mean and variance results of the j^{th} user. These mean and variance values are obtained by excluding the j^{th} user. The result in (15) and (17) determines the impact of not including each cooperative user during the average mean and variance observation measurement. As all MUs including AY, AN, AO, and RO have dissimilar results of the mean and variance in comparison with normal SUs, therefore the average results attained against these users are different from the normal SUs in (15) and (17).

The KLD value for the j^{th} SU is determined between the individual sensing results in (10), (12), and the information provided by all other SU information as in (15) and (17) as

$$K_j(i) = KL\left(\mu_{ja'}(i), \mu_{ja'}(i), \sigma_{jb'}^2(i), \sigma_{jb}^2(i)\right) \tag{18}$$

where $K_j(i)$ denotes the KLD result in the presence and absence hypothesis of the j^{th} SU in the i^{th} interval. These KLD scores against each SU sensing are modified as

$$c_j(i) = \left(\frac{1}{K_j(i)}\right), \quad i \in 1, \dots, N,\ j \in 1, \dots, M \tag{19}$$

The result in (19) is normalized for assigning weights to each SU decision as

$$w_j(i) = \left(\frac{c_j(i)}{\sum_{j=1}^{M}c_j(i)}\right), \quad i \in 1, \dots, N,\ j \in 1, \dots, M \tag{20}$$

In (20) the users with abnormal behavior acquire lower weights in comparison with normal users.

Table 1 shows the weight measurement for the normal and malicious users against various signals-to-noise ratios. These weights are obtained for the case when one of the four categories of MUs participates in CSS. In Table 1 as the value of signal-to-noise ratio increases the weight assigned to these MUs decreases while the normal user's weights increase.

Similarly, Table 2 shows the weights for the case when all four categories of MUs participate in CSS. In Table 2, the weight result assigned to each MU is shown along with the average weights received by all the normal cooperative SUs. In this case, the different weights received by these MUs approach zero with increasing signal-to-noise ratio while the normal SUs weights increase with increasing signal-to-noise ratio.

Table 1: KLD weights assigned by the FC under one category of MU participation.

SNR (dB)	Weights				
	AY only	AN only	AO only	RO only	Normal User
-20	0.006757	0.006553	0.016615	0.008775	0.080399
-19	0.006750	0.006551	0.008679	0.006123	0.080798
-18	0.006745	0.006547	0.008341	0.005763	0.081049
-17	0.006740	0.006544	0.008341	0.005757	0.081110
-16	0.006737	0.006539	0.006206	0.005616	0.081198
-15	0.006731	0.006537	0.006186	0.005537	0.081231
-14	0.006722	0.006532	0.006164	0.005393	0.081266
-13	0.006717	0.006530	0.005722	0.005295	0.081306
-12	0.006715	0.006526	0.005722	0.005290	0.081324
-11	0.006711	0.006525	0.005629	0.004688	0.081428
-10	0.006709	0.006521	0.005629	0.004318	0.081441
-9	0.006706	0.006518	0.005190	0.003863	0.081545
-8	0.006704	0.006516	0.004947	0.003836	0.081636
-7	0.006701	0.006510	0.003674	0.003773	0.081739
-6	0.006692	0.006505	0.001509	0.003198	0.081777
-5	0.006687	0.006502	0.001507	0.001335	0.082069

Table 2: KLD weights assigned by the FC when all categories of MUs participate.

SNR (dB)	Weights				
	1 AY	1 AN	1 AO	1 RO	Normal User
-20	0.000682	0.000359	0.001661	0.065425	0.077865
-19	0.000523	0.000331	0.001155	0.012339	0.082344
-18	0.000466	0.000319	0.001085	0.006149	0.082800
-17	0.000379	0.000277	0.001037	0.005841	0.082875
-16	0.000287	0.000212	0.000825	0.005060	0.082967
-15	0.000229	0.000169	0.000817	0.004495	0.082984
-14	0.000175	0.000159	0.000766	0.004449	0.083008
-13	0.000160	0.000139	0.000645	0.004355	0.083035
-12	0.000137	0.000113	0.000637	0.003774	0.083047
-11	0.000112	0.000080	0.000477	0.002980	0.083048
-10	0.000096	0.000079	0.000469	0.002719	0.083058
-9	0.000095	0.000070	0.000285	0.002563	0.083061
-8	0.000094	0.000069	0.000242	0.002524	0.083136
-7	0.000082	0.000066	0.000222	0.002486	0.083254
-6	0.000055	0.000039	0.000137	0.001171	0.083307
-5	0.000010	0.000008	0.000057	0.000266	0.083694

3.3. Global Decision at FC. On the basis of weighted results measured to guarantee the authenticity of each SU sensing information in (20), the global statement $G_B(i)$ is declared by the FC as

$$G_B(i) = \left\{ \begin{array}{l} H_1, \ \sum_{j=1}^{M} \boldsymbol{w}_j(i) * E_j(i) \geq \varepsilon \\ H_0, \ otherwise \end{array} \right\}, \quad i \in 1, \ldots, N \tag{21}$$

where $\boldsymbol{w}_j$ is the weight assigned to the j^{th} user energy in the data fusion at the FC and ε the threshold value for the detection of the PU. The lesser weight results are charged by the FC against the sensing information of a user with malicious behavior, while the normal user sensing report is assigned with a higher weight value. All MUs including AY, AN, AO, and RO are easily identified by the proposed scheme with their KLD behavior. The normal SUs have a higher KLD result because they have less inconsistency with the average of all other users sensing information. The MUs receive minimum weight because the information provided by MUs deviates more from the average sensing information provided by all other SUs. It is therefore noticeable that these MUs get lower weights as compared with normal SUs.

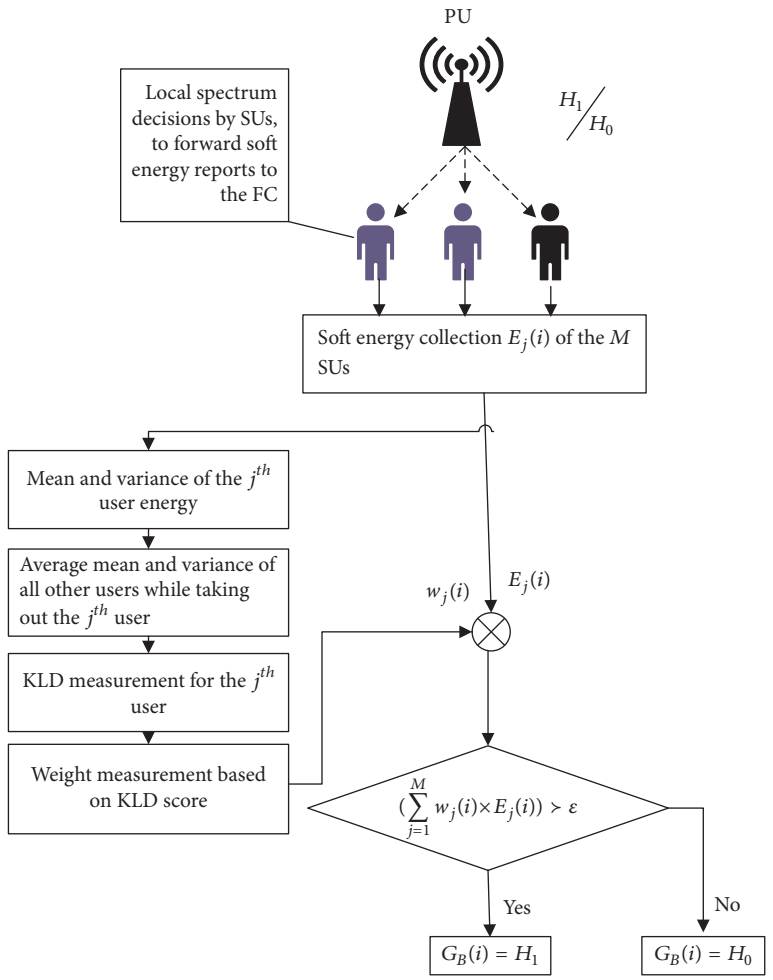


Figure 3: Flowchart of the proposed CSS mechanism.

3.4. Updating Statistics Based on the Global Decision. Due to the nonavailability of the exact information about the PU, optimal values of the means (μ_{j1}, μ_{j0}) and variances $(\sigma_{j0}^2, \sigma_{j1}^2)$ for measuring KLD are not possible. It is therefore good to consider the global decision $G_B(i)$ results as an estimate of the primary signal, in order to calculate and update these values. The updated mean and variance will be used by the FC in the KLD calculation in the next sensing interval.

$$E_{j1} = \left\{E_j(i) \mid H_1\right\} \approx \left\{E_j(i) \mid G_B(i) = H_1\right\} \quad (22)$$

$$E_{j0} = \left\{E_j(i) \mid H_0\right\} \approx \left\{E_j(i) \mid G_B(i) = H_0\right\} \quad (23)$$

After the establishment of universal decision at the FC, resultant decision $G_B(i) = 1$ will update mean μ_{j1} and variance σ_{j1}^2 values under the H_1 hypothesis as follows:

$$\begin{aligned} \mu_{j1} &= B_1 \mu_{j1} + B_2 Z_j(i) \\ \sigma_{j1}^2 &= B_1 \sigma_{j1}^2 + \frac{B_1}{B_2}\left[Z_j(i) - \mu_{j1}\right]^2 \end{aligned} \quad (24)$$

Similarly, the decision $G_B(i) = 0$ will update mean μ_{j0} and variance σ_{j0}^2 under the H_0 hypothesis for all cooperative users as

$$\begin{aligned} \mu_{j0} &= B_1 \mu_{j0} + B_2 E_j(i) \\ \sigma_{j0}^2 &= B_1 \sigma_{j0}^2 + \frac{B_1}{B_2}\left[E_j(i) - \mu_{j0}\right]^2 \end{aligned} \quad (25)$$

In (25) $B_1 = z/(z-1)$ and $B_2 = 1/z$, where z indicate the window size of the sensing history for the estimated mean and variance.

The proposed scheme flowchart diagram in Figure 3 illustrates the stepwise procedure of the local detection, KL divergence measurement based on the weight assignments at the FC, and global decision establishment by the FC.

4. Numerical Results and Discussion

In order to get simulation results for the CRN, parameters settings are made with 10, 16, 20, and 30 total cooperative users. Out of the total cooperative SUs, four users are intentionally selected as AY, AO, RO, and AO nature of MUs. The average signal-to-noise ratios for the simulation

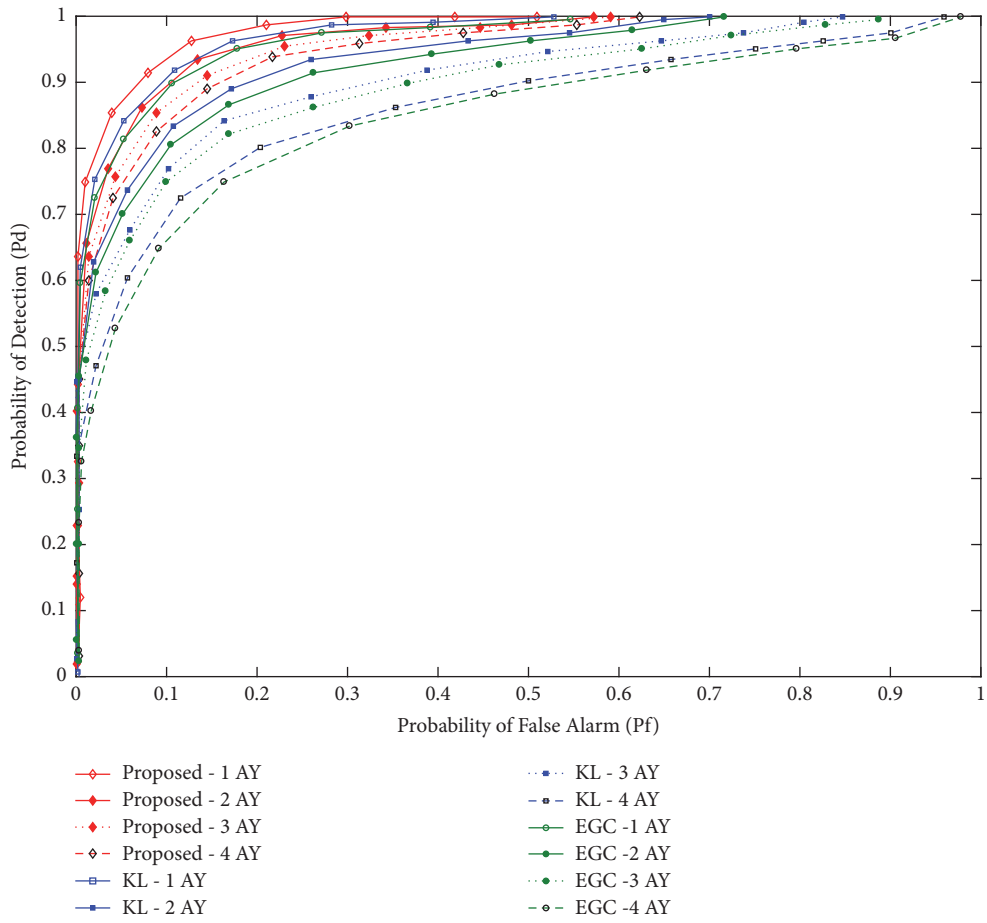


Figure 4: Detection vs. false alarm results with AY malicious users.

are selected as -20 dB to -5 dB for all SUs. The sensing time for each SU is selected as 1 ms containing 270 samples in each sensing interval. Total sensing intervals for the cooperative users are selected as 200. The RO users perform malicious acts probabilistically in the intervals 1 to N. The window size (z) for updating mean and variance is selected as 270. In the study, all 4 categories of MUs, i.e., AY, AN, AO, and RO, are spread evenly.

The proposed KLD performance is compared with traditional KL and equal gain combination (EGC) schemes in 6 different cases as follows.

Case 1. In this case ROC results are drawn between the proposed method and traditional KL and EGC scheme under various signal-to-noise ratio values of -20 dB to -5 dB as displayed in Figure 4. MUs are selected as AY only in the first part of the comparison in Figure 4. Results are obtained for all combinations by taking the total AY users number as 1, 2, 3, and 4 subsequently. The results show that the proposed KLD scheme is more secure against the increasing number of AY users 1 to 4 and has better detection probability results in comparison with all other schemes. In Figure 4 when there is only 1 AY user active in CSS the ROC results of all fusion schemes are less affected, but when the total number of AY users is increased to 3 and 4 the proposed KL results dominate the traditional KL and EGC schemes by producing a high detection with fewer false alarms. The EGC scheme is more affected by the increasing number of AY users because EGC is giving equal weight to the detection performance of normal and AY users. The proposed KL is able to assign less weight to the AY users in comparison with normal SUs as it is clear from the average weight value measured against each AY users in Table 1. The less weight assigned to the AY users reduces the false data effect of the AY users participation in CSS. The harmful effect of the AY users contribution in CSS is further reduced with increasing average SNR by lowering the weight assignment to them in the global decision.

Case 2. In this part of the simulation, all parameters are kept similar to Case 1 with changing only the nature of MUs from AY to AN user. Comparison is made between proposed KL, traditional KL, and EGC scheme by testing the system against increasing AY users number from 1 to 4 as in Figure 5.

Since the proposed KL is treating AY and AN users similarly in determining the KLD, therefore using proposed KLD the weight that AN user receives is almost equal to the AY user weights in Case 1. The ROC performance of the proposed and all other schemes against the AN scenario is very similar to Case 1, due to the similar behavior of the AN user to that of the AY user. As in Case 1, when the numbers

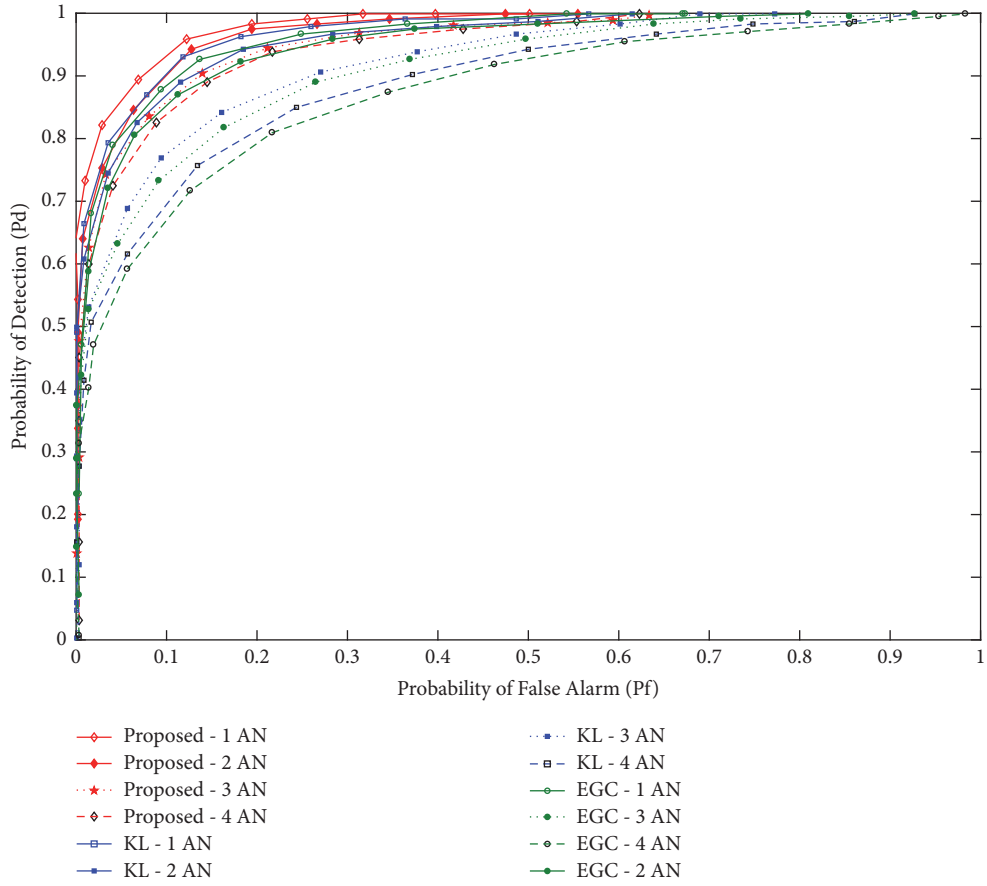


Figure 5: Detection vs. false alarm results with AN malicious users.

of AN users are increased from 1 to 4, the proposed KL is less affected by this increment in Figure 5. All AN users receive lower weight due to their KLD score results while the normal SUs receive higher weights in comparison with AN users which results in better performance of the proposed KL scheme. The traditional KL and EGC schemes performance in detecting the licensed PU channel reduces more quickly in comparison with the proposed KL method as the total AY increases from 1 to 4. The gap in the ROC curves of the traditional fusion schemes becomes wider for a total of 4 AN users from the one when only 1 AN user takes part as shown in Figure 5.

Case 3. In the third scenario detection and false alarm results are obtained for the increasing number of AO users from 1 to 4 in Figure 6 with the same parameters in Cases 1 and 2. Since the AO users have their mean and variance results opposite to the average mean and variance values provided by all other users, therefore, the proposed KL method is able to generate lower reliability report in terms of weight for the AO user in comparison with normal cooperative users. The results show that as the number of AO users increases to 4 few drops are observed in the ROC curve of the proposed scheme as compared with the traditional KL and EGC scheme. In comparison with Cases 1 and 2, the traditional soft combination schemes like KL and EGC performance degrade even more. The existence of AO users results in less correct detection and high false alarm rate of the PU spectrum for the EGC and KL scheme. Proposed method results in Figure 6 are followed by the KL while EGC has shown its worst performance among all fusions.

Case 4. The ROC results for the scenario in which only RO user participates in CSS are depicted in Figure 7. The RO user hides its malicious identity by acting probabilistically as AO at randomly selected sensing intervals in the N total intervals and is difficult to catch with the provided statistics.

The traditional KL and EGC schemes are not able to handle the RO user information intelligently and their ROC results degrade severely with the increased number of RO participations in Figure 7.

The proposed KL scheme is able to identify the RO users when they perform malicious acts probabilistically and generate better detection and false alarm results in Figure 7 compared with the traditional KL and EGC schemes. Results show that the proposed KLD is less affected by the increasing number of RO users, unlike the traditional EGC and KLD schemes. All the RO nature users in the proposed CSS receive lesser weights in comparison with weights obtained by the normal SUs because their malicious behavior is easily caught by the proposed KL scheme.

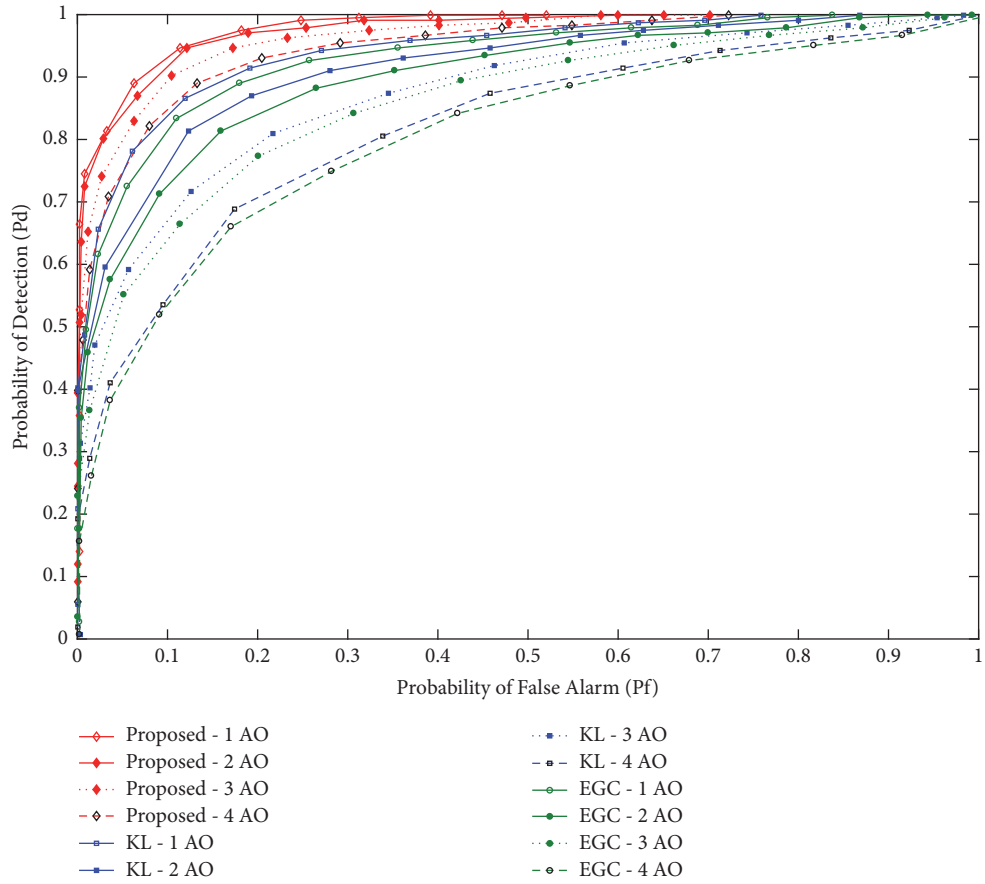


Figure 6: Detection vs. false alarm results with AO malicious users.

Case 5. In this part of the simulation as in Figure 8 MUs are equally selected in numbers as AY, AN, AO, and RO categories. The simulation is performed against an average signal-to-noise ratio of -12.5 dB.

The detection and false alarm probability results are obtained for the proposed, traditional KL, and EGC schemes for a total of 10, 20, and 30 cooperative SUs in Figure 8 under average signal-to-noise ratio of -12.5 dB and total 4 MUs. Figure 8 shows that the one-to-many relation based KLD scheme has better ROC performance than all other schemes with different levels of the cooperative users. It is noticeable that the detection performance is enhanced for all combination schemes with the increasing number of total cooperative and fixed MUs. The proposed method ROC results are more precise and superior to the traditional schemes, i.e., KLD and EGC schemes, at all levels of the total sensing users.

Case 6. In this case, the AY, AO, AN, and RO users number is kept the same. The total number of participating SUs in CSS is kept fixed as 16 and different ROC results are plotted for the one-to-many relations based KLD and other soft combination schemes at different levels of the averages signal-to-noise ratios.

The simulation results in Figure 9 show that under fixed malicious and total cooperative users the ROC performance rises with increasing signal-to-noise ratio values for all combination schemes. Similarly, in Figure 9 as the signal-to-noise ratio value increases from -15.5 dB to -9.5 dB, all schemes are able to generate a high detection rate with minimum false alarm. The proposed scheme ROC results are more accurate and precise than the traditional combination schemes at both SNR levels. The proposed method ROC improvement with increasing signal-to-noise ratio is due to more clear distinction in the energy distribution of the absence and presence hypothesis information provided by the normal and MUs. As the SNR increases in Figure 9, the proposed method detection results rise more quickly against other methods. These results also show that the CSS performance improves more with the increasing signal-to-noise ratio information in Case 6 as compared with the increasing total number of users in Case 5.

All the above experimental results clarify the fact that by following the proposed one-to-many relations based KLD method an improvement is obvious in the sensing performance at the FC. This improvement in performance is achieved by raising the detection probability and lowering the false alarm results leading to a reduction in the error probability of the system. The proposed fusion combination scheme shows optimum and accurate results in the presence of MUs. The use of the proposed method for calculating

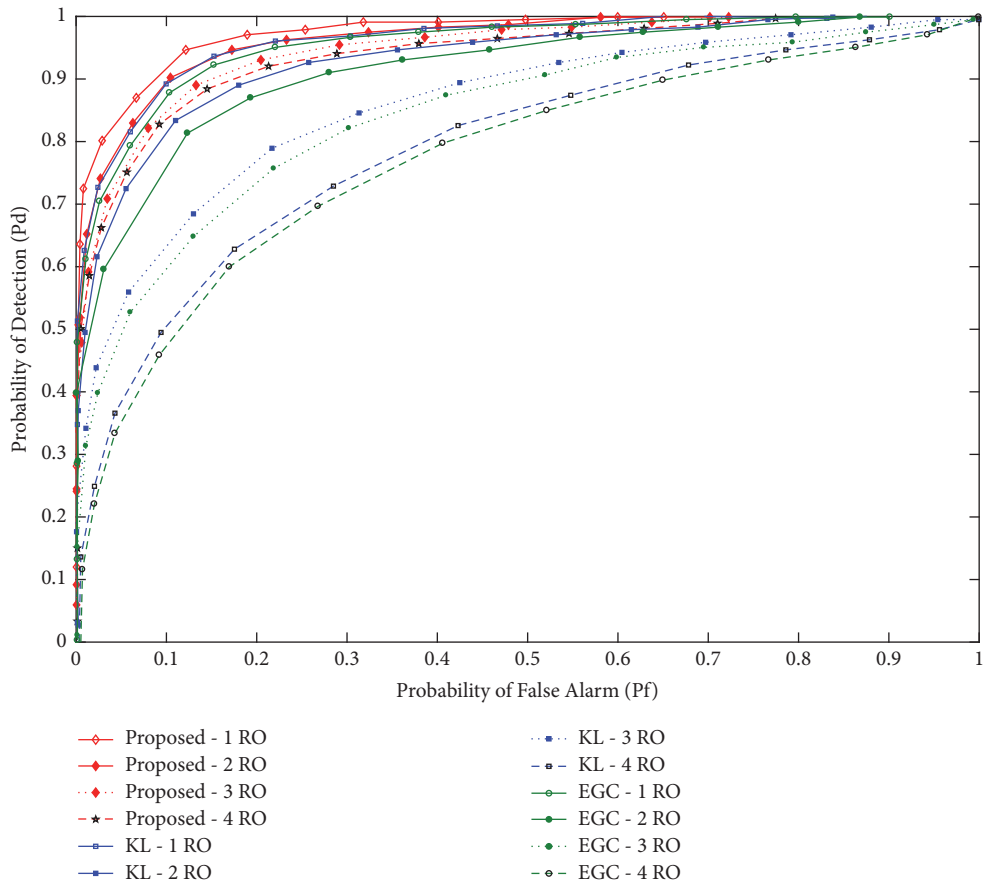


Figure 7: Detection vs. false alarm results with RO users.

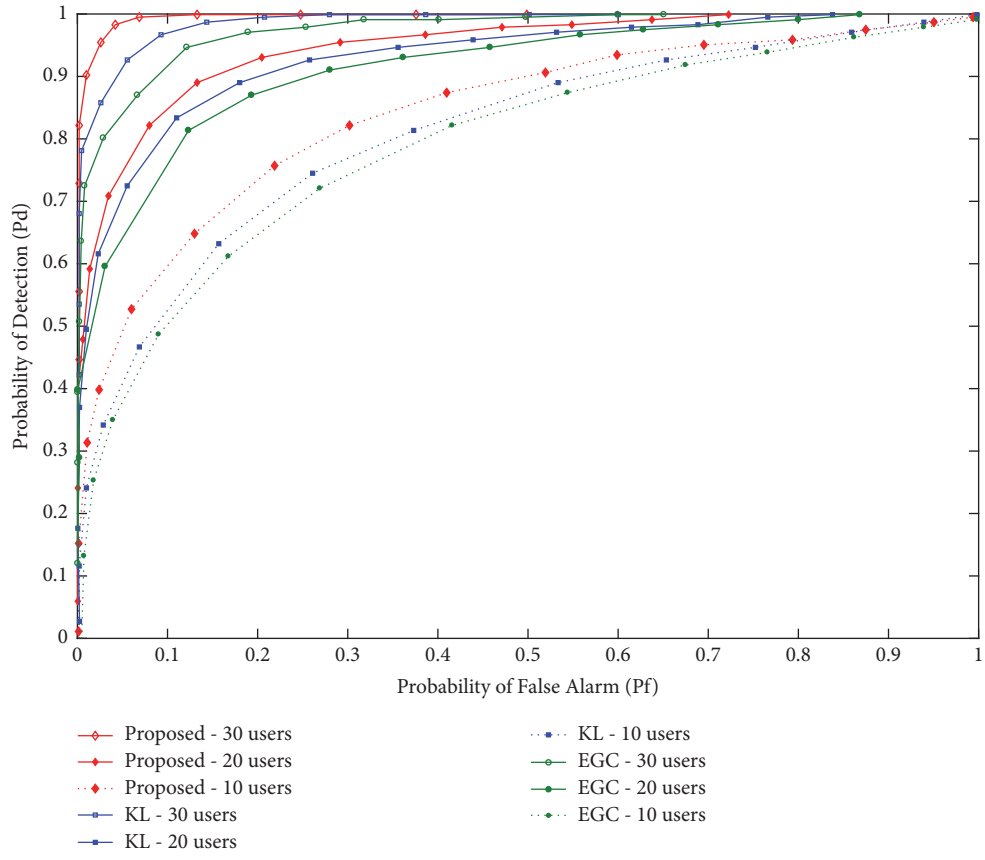


Figure 8: Detection vs. false alarm results with all MUs and different number of total reporting users.

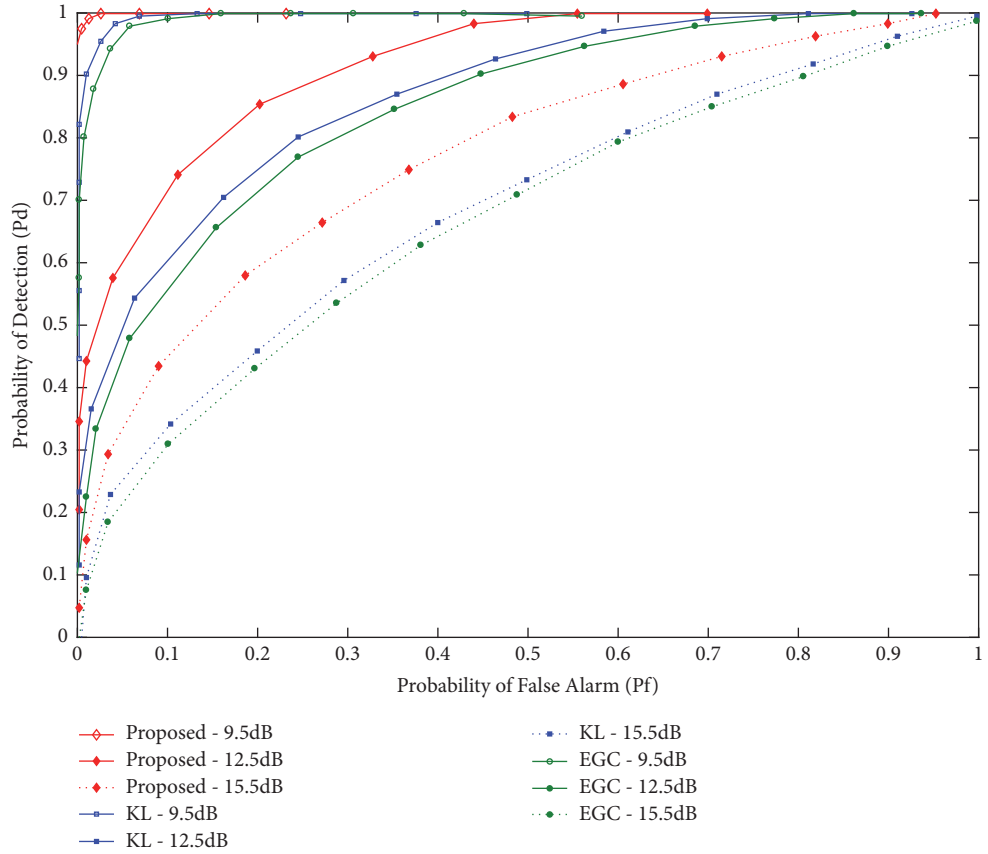


Figure 9: Detection vs. false alarm results with all MUs and different levels of signal-to-noise ratios.

weights following soft combination scheme makes the proposed CSS results more valid in the presence of malicious user. The simulation results show that the risk of AY, AN, RO, and AO users with CSS significantly reduces by adopting the proposed scheme. It is clear from the graphical result that the process of cooperation turns out to be more precise by using the suggested methodology. The one-to-many relation based KLD is able to generate better sensing results, by assigning lower weights to the MUs information, and is able to eliminate the effect of MUs in the resultant CSS.

5. Conclusion

In this paper, the efficiency degradation due to the presence of abnormal users in CSS is minimized using one-to-many relationship based KLD method for the PU detection. Functionality of the proposed scheme is verified in the presence of AY, AN, AO, and RO type MUs. FC first receives the individual sensing information of all SUs and then applies the proposed method for measuring weights against each SU. MUs with abnormal behavior as compared with normal SUs are given lower weights by the proposed scheme, while the normal SUs receive higher weights. FC further employs these weights in combining the sensing information of all SUs in predicting a global decision. The results show that the user with abnormal behavior has less impact on the global decision as compared to a normal SU decision. Simulation result reflects the superiority and authenticity of the proposed scheme in producing more precise and reliable decisions as compared with EGC and traditional KL fusion schemes.

Data Availability

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

Disclosure

The authors received no specific funding for this work.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, 2008.

[2] R. Chen, J.-M. Park, Y. T. Hou, and J. H. Reed, "Toward secure distributed spectrum sensing in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 50–55, 2008.

[3] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proceedings of the IEEE International Conference on Communications*, pp. 1658–1663, Istanbul, Turkey, July 2006.

[4] S. Kaur, "Intelligence in wireless networks with cognitive radio networks!," *IETE Technical Review*, vol. 30, no. 1, pp. 6–11, 2013.

[5] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: state-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[6] L. Zhai, H. Wang, and C. Gao, "A spectrum access based on quality of service in cognitive radio networks," *PLOS ONE*, vol. 11, no. 5, pp. 2005–2009, 2016.

[7] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, 2009.

[8] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 814–823, 2010.

[9] C.-H. Hwang, G.-L. Lai, and S.-C. Chen, "Spectrum sensing in wideband OFDM based cognitive radio," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 709–719, 2015.

[10] A. Ali and W. Hamouda, "Spectrum monitoring using energy ratio algorithm for OFDM-based cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 2257–2268, 2015.

[11] L. Miao, Z. Sun, and Z. Jie, "The Parallel Algorithm Based on Genetic Algorithm for Improving the Performance of Cognitive Radio," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 5986482, 6 pages, 2018.

[12] A. Elahi, I. M. Qureshi, F. Zaman, N. Gul, and T. Saleem, "Suppression of mutual interference in noncontiguous orthogonal frequency division multiplexing based cognitive radio systems," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

[13] M. A. Raja, M. S. Aslam, N. I. Chaudhary, and W. U. Khan, "Bio-inspired heuristics hybrid with interior-point method for active noise control systems without identification of secondary path," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 2, pp. 246–259, 2018.

[14] Y. Eghbali, H. Hassani, A. Koohian, and M. Ahmadian-Attari, "Improved energy detector for wideband spectrum sensing in cognitive radio networks," *Radioengineering*, vol. 23, no. 1, pp. 430–434, 2014.

[15] H. Guo, W. Jiang, and W. Luo, "Linear Soft Combination for Cooperative Spectrum Sensing in Cognitive Radio Networks," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1573–1576, 2017.

[16] J. Wang, S. Feng, Q. Wu, X. Zheng, Y. Xu, and G. Ding, "A robust cooperative spectrum sensing scheme based on Dempster-Shafer theory and trustworthiness degree calculation in cognitive radio networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, 2014.

[17] J. Tong, M. Jin, Q. Guo, and Y. Li, "Cooperative Spectrum Sensing: A Blind and Soft Fusion Detector," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2726–2737, 2018.

[18] M. Zheng, W. Liang, H. Yu, and M. Song, "SMCSS: A Quick and Reliable Cooperative Spectrum Sensing Scheme for Cognitive Industrial Wireless Networks," *IEEE Access*, vol. 4, pp. 9308–9319, 2016.

[19] J. Wang, I. Chen, J. J. Tsai, and D. Wang, "Trust-based mechanism design for cooperative spectrum sensing in cognitive radio networks," *Computer Communications*, vol. 116, pp. 90–100, 2018.

[20] F. Zaman, "Joint Angle-Amplitude Estimation for Multiple Signals with L-Structured Arrays Using Bioinspired Computing," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9428196, 12 pages, 2017.

[21] X. Liu, J. Yan, and K. Chen, "Optimal energy harvest-based weighed cooperative spectrum sensing in cognitive radio," in *Proceedings of the 2016 International Conference on Computing, Networking and Communications (ICNC)*, pp. 1–5, Kauai, HI, USA, Feburary 2016.

[22] P. D. Sutton, K. E. Nolan, and L. E. Doyle, "Cyclostationary signatures in practical cognitive radio applications," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 13–24, 2008.

[23] S. A. Hosseini, B. Abolhassani, and S. M. S. Sadough, "A new protocol for cooperative spectrum sharing in mobile cognitive radio networks," *Radioengineering*, vol. 24, no. 3, pp. 757–764, 2015.

[24] L. Zhang, G. Ding, Q. Wu, and F. Song, "Defending against byzantine attack in cooperative spectrum sensing: defense reference and performance analysis," *IEEE Access*, vol. 4, pp. 4011–4024, 2016.

[25] F. Penna, Y. Sun, L. Dolecek, and D. Cabric, "Detecting and counteracting statistical attacks in cooperative spectrum sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1806–1822, 2012.

[26] K. Zeng, P. Pawełczak, and D. Čabrić, "Reputation-based cooperative spectrum sensing with trusted nodes assistance," *IEEE Communications Letters*, vol. 14, no. 3, pp. 226–228, 2010.

[27] R. Yu, Y. Zhang, Y. Liu, S. Gjessing, and M. Guizani, "Securing cognitive radio networks against primary user emulation attacks," *IEEE Network*, vol. 29, no. 4, pp. 68–74, 2015.

[28] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," *IEEE Transactions on Wireless Communications*, vol. 9, no. 8, pp. 2488–2497, 2010.

[29] V. Heip and I. Koo, "A sequential cooperative spectrum sensing scheme based on cognitive user reputation," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1147–1152, 2012.

[30] H. Guo, N. Reisi, W. Jiang, and W. Luo, "Soft combination for cooperative spectrum sensing in fading channels," *IEEE Access*, vol. 5, pp. 975–986, 2017.

[31] M. Emami, H. Zarrabi, M. R. Jabbarpour, M. Sadat Taheri, and J. J. Jung, "A soft cooperative spectrum sensing in the presence of most destructive smart PUEA using energy detector," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 15, p. e4524, 2018.

[32] Y. L. Lee, W. K. Saad, A. A. El-Saleh, and M. Ismail, "Improved detection performance of cognitive radio networks in AWGN and rayleigh fading environments," *Journal of Applied Research and Technology*, vol. 11, no. 3, pp. 437–446, 2013.

[33] D. Teguig, B. Scheers, and V. Le Nir, "Data fusion schemes for cooperative spectrum sensing in cognitive radio networks," in *Proceedings of the 2012 Military Communications and Information Systems Conference, MCC 2012*, pp. 104–110, Poland, October 2012.

[34] J. So and W. Sung, "Group-Based Multibit Cooperative Spectrum Sensing for Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10193–10198, 2016.

[35] S. Nallagonda, Y. R. Kumar, and P. Shilpa, "Analysis of hard-decision and soft-data fusion schemes for cooperative spectrum sensing in rayleigh fading channel," in *Proceedings of the*

7th IEEE International Advanced Computing Conference, IACC 2017, pp. 220–225, India, January 2017.

[36] W. Ejaz, G. Hattab, T. Attia, M. Ibnkahla, F. Abdelkefi, and M. Siala, "Joint Quantization and Confidence-Based Generalized Combining Scheme for Cooperative Spectrum Sensing," *IEEE Systems Journal*, vol. 12, no. 2, pp. 1909–1920, 2018.

[37] H. Wu, F. Yao, Y. Chen, Y. Liu, and T. Liang, "Multibit-Quantization-Based Collaborative Spectrum Sensing Scheme for Cognitive Sensor Networks," *IEEE Access*, vol. 5, pp. 25207–25216, 2017.

[38] M. A. Raja, A. A. Shah, A. Mehmood, N. I. Chaudhary, and M. S. Aslam, "Bio-inspired computational heuristics for parameter estimation of nonlinear Hammerstein controlled autoregressive system," *Neural Computing and Applications*, vol. 29, no. 12, pp. 1455–1474, 2018.

[39] S. Bhattacharjee, P. Das, S. Mandal, and B. Sardar, "Optimization of probability of false alarm and probability of detection in cognitive radio networks using GA," in *Proceedings of the 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, pp. 53–57, Kolkata, India, July 2015.

[40] N. Gul, I. M. Qureshi, A. Elahi, and I. Rasool, "Defense against Malicious Users in Cooperative Spectrum Sensing Using Genetic Algorithm," *International Journal of Antennas and Propagation*, vol. 2018, 2018.

[41] N. Gul, A. Naveed, A. Elahi, T. Saleemkhattak, and I. M. Qureshi, "A combination of double sided neighbor distance and Genetic Algorithm in cooperative spectrum sensing against malicious users," in *Proceedings of the 14th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2017*, pp. 746–753, Pakistan, January 2017.

[42] U. Mehboob, J. Qadir, S. Ali, and A. Vasilakos, "Genetic algorithms in wireless networking: techniques, applications, and issues," *Soft Computing*, vol. 20, no. 6, pp. 2467–2501, 2016.

[43] M. Akbari and M. Ghanbarisabagh, "A novel evolutionary-based cooperative spectrum sensing mechanism for cognitive radio networks," *Wireless Personal Communications*, vol. 79, no. 2, pp. 1017–1030, 2014.

[44] A. A. EL-Saleh and K. Hussain, "Cognitive radio engine model utilizing soft fusion based genetic algorithm for cooperative spectrum optimization," *International Journal of Computer Networks and Communications*, vol. 2, no. 3, pp. 169–173, 2013.

[45] M. Taha and D. Alnadi, "Threshold adaptation in spectrum sensing for cognitive radio using particle swarm optimization," in *Proceedings of the International Conference on Control, Engineering & Information Technology*, pp. 223–228, Sousse, Tunisia, 2014.

[46] N. Gul, I. M. Qureshi, A. Omar, A. Elahi, S. Khan, and B. Podobnik, "History based forward and feedback mechanism in cooperative spectrum sensing including malicious users in cognitive radio network," *PLoS ONE*, vol. 12, no. 8, p. e0183387, 2017.

[47] H. Vu-Van and I. Koo, "A robust cooperative spectrum sensing based on kullback-leibler divergence," *IEICE Transactions on Communications*, vol. E95-B, no. 4, pp. 1286–1290, 2012.