Zuber Patel
Shilpi Gupta (Eds.)

220

Future Internet Technologies and Trends

LNICST

First International Conference, ICFITT 2017
Surat, India, August 31 – September 2, 2017
Proceedings

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 220

More information about this series at http://www.springer.com/series/8197

Zuber Patel · Shilpi Gupta (Eds.)

Future Internet Technologies and Trends

First International Conference, ICFITT 2017
Surat, India, August 31 – September 2, 2017
Proceedings

Springer

Editors
Zuber Patel
Sardar Vallabhbhai National Institute of Technology
Surat
India

Shilpi Gupta
Sardar Vallabhbhai National Institute of Technology
Surat
India

ISSN 1867-8211 ISSN 1867-822X (electronic)
Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering
ISBN 978-3-319-73711-9 ISBN 978-3-319-73712-6 (eBook)
https://doi.org/10.1007/978-3-319-73712-6

Library of Congress Control Number: 2017963761

Printed on acid-free paper

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

Preface

It gives us immense pleasure to present the proceeding of the First EAI International Conference on Future Internet Technologies and Trends (ICFITT 2017). The conference was organized by the Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat (Gujarat State, India).

Holding an international conference requires good team work, total dedication, and devotion of time. It was very nice to join hands with the European Alliance for Innovation (EAI), which not only provided all the support for maintaining the website and the Confy system (the conference management application for paper uploading, reviewing, registration etc.) but also provided partial financial support as well as support for having the proceedings published by Springer.

There cannot be boundaries for knowledge. A conference is a place where experts, researchers, academics and industry professionals with various experiences and proficiencies gather to discuss the state of the art. The theme of our conference was next-generation requirements for extremely high speed data communications, IoT, and security, i.e., mainly 5G requirements, which is reflected in the title of the conference. The proposed topics were 4G technologies—LTE and various aspects; broadband technology; cognitive radio; vehicular technology; gigabit wireless networks; radio over fiber; IOT—protocols; architecture; various technologies adopted in IOT; sensors; actuators and new consumer devices/things; fundamental aspects of LTE-A systems; wireless technologies for IoT; ubiquitous computing; data management and big data; security aspects etc.

The journey of our conference started with a past student of the EC Department at SVNIT, Mr. Yatindra Shashi, who was working under Dr. Imrich Chlamtac, the President of EAI. Dr. Imrich agreed to take on the role of chair of the Steering Committee and Dr. Upena Dalal took on the role of general chair of the Organizing Committee. We take this opportunity to thank Dr. Imrich for his involvement and commitment. Various committees were then formed and started on the journey almost one year ago. During this journey, whole-hearted support was received from various EAI members as well as from the staff of the Electronics Engineering Department at SVNIT. In total, 66 papers were received and after a plagiarism check as well as a rigorous review, 45 papers were selected for presentation.

We thank our esteemed sponsors the EAI, Keysight Technologies, Gujcost, DRDO, SVNIT Alumni Association Surat Chapter (and ErBhadresh Shah of SwapnPriti foundations), and Royal Electronics, without whose financial support this conference would not have taken place. There was tremendous support from the Technical Program Committee team and reviewers, all the keynote speakers, invited paper speakers, session chairs, and registered participants. We would like to express our deep gratitude the director, registrar, chair CCE, deputy registrar (A/C), and administrative staff for providing the infrastructural support and aiding in various arrangements.

December 2017

Zuber Patel
Shilpi Gupta

Organization

Steering Committee

Steering Committee Chair

Imrich Chlamtac	CREATE-NET and University of Trento, Italy

Steering Committee Member

Upena Dalal	Sardar Vallabhbhai National Institute of Technology, India

Organizing Committee

General Chair

Upena Dalal	Sardar Vallabhbhai National Institute of Technology, India

General Co-chairs

Jignesh Sarvaiya	SVNIT, Surat, India
Shweta Shah	SVNIT, Surat, India

Technical Program Committee Chairs

Anand Darji	SVNIT, Surat, India
Rasika Dhavse	SVNIT, Surat, India

Workshops Chairs

Niteen Patel	SCET, Surat, India
Piyush Patel	SVNIT, Surat, India

Publicity and Social Media Chairs

Pinal Engineer	SVNIT, Surat, India
Yatindra Shashi	ICT Innovation, TU Berlin/UNITN, Germany

Sponsorship and Exhibits Chair

Maulin Joshi	SCET, Surat, India

Publications Chairs

Shilpi Gupta	SVNIT, Surat, India
Zuber Patel	SVNIT, Surat, India

Local Chair

Jigisha Patel	SVNIT, Surat, India

Web Chair

Ramesh Solanki	SVNIT, Surat, India

Conference Manager

Monika Szabova	European Alliance for Innovation (EAI)

Technical Program Committee

Pradip Mainali	TP Division, Technologiepark, Zwijnaarde, Belgium
Chintan Bhatt	CHRUSAT, Changa, India
Pavel Loskot	College of Engineering, Swansea University, UK
Rakesh Jha	SMVD University, Katra (J&K), India
Vishal Wankhede	SNJBs K. B. Jain College of Engineering, Chandwad, India
Chirag Paunwala	SCET, Surat, India
Jay Joshi	Shri S'ad Vidya Mandal Institute of Technology, Bharuch, India
J. Nirmal	K. J. Somaiya College of Engineering, Mumbai, India
Robin Singh Bhadoria	Indian Institute of Technology Indore, India
Ganesh Deka	Tura, DGT, Ministry of Skill Development and Entrepreneupship, Tura Meghalaya, India
K. M. Sunjiv Soyjaudah	University of Mauritius, Reduit, Mauritius
Y. P. Kosta	Marwadi Education Foundation, Rajkot, India
Tanmay Pawar	BVM Engineering College, Vallabh Vidyanagar, Anand, India
Nishith Bhatt	DesignTech Systems Ltd., India
Manik Sharma	DAV University, India
P. K. Shah	SVNIT, Surat, India
Abhilash Mandloi	SVNIT, Surat, India

Contents

LOGO: A New Distributed Leader Election Algorithm in WSNs with Low Energy Consumption

Ahcène Bounceur[1(✉)], Madani Bezoui[2], Umber Noreen[1], Reinhardt Euler[1], Farid Lalem[1], Mohammad Hammoudeh[3], and Sohail Jabbar[4]

[1] Université de Bretagne Occidentale, CNRS Lab-STICC Laboratory, UMR 6285, Brest, France
Ahcene.Bounceur@univ-brest.fr
[2] Université de Boumerdes, Boumerdes, Algeria
[3] University of Manchester, Manchester, UK
[4] Department of Computer Science, National Textile University, Faisalabad, Pakistan

Abstract. The Leader Election Algorithm is used to select a specific node in distributed systems. In the case of Wireless Sensor Networks, this node can be the one having the maximum energy, the one situated on the extreme left in a given area or the one having the maximum identifier. A node situated on the extreme left, for instance, can be used to find the boundary nodes of a network embedded in the plane. The classical algorithm allowing to find such a node is called the Minimum Finding Algorithm. In this algorithm, each node sends its value in a broadcast mode each time a better value is received. This process is very energy consuming and not reliable since it may be subject to an important number of collisions and lost messages. In this paper, we propose a new algorithm called LOGO (Local Optima to Global Optimum) where some local leaders will send a message to a given node, which will designate the global leader. This process is more reliable since broadcast messages are sent only twice by each node, and the other communications are based on a direct sending. The obtained results show that the proposed algorithm reduces the energy consumption with rates that can exceed 95% compared with the classical Minimum Finding Algorithm.

Keywords: Wireless sensor network · Leader election
Distributed algorithms

1 Introduction and Related Work

Wireless Sensor Networks (WSNs) are useful in situations where we need to measure environmental data, especially when the measures must be taken in dangerous or inaccessible places. In the context of Internet of Things (IoTs) and Smart-Cities, WSNs can be used to detect free places in car parkings, to

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 1–16, 2018.
https://doi.org/10.1007/978-3-319-73712-6_1

secure and detect intrusions around sensitive sites, to predict and detect fires, etc. They are composed of autonomous sensor nodes that communicate between them using short wireless communication in order to exchange messages and data. They can also communicate with a static or mobile base station in order to transmit the collected data.

The main context of this paper, is the surveillance of sensitive and dangerous sites where one needs to find the boundary nodes of a WSN. A recent algorithm, called D-LPCN [1] (Distributed Least Polar-angle Connected Node), can be used for this purpose. The nodes must communicate with each other. This characteristic is necessary in order to be able to detect faulty nodes [2]. These algorithms start from the node which is on the extreme left of the network. To find this particular node, one can use any Leader Election algorithm which can also be used for other applications and actions, like for example, coordination, cooperation, etc. Leader election is a complex problem in distributed systems since the data are distributed among the different nodes, which are geographically separated as is the case for WSNs. Several approaches are available to deal with this problem. The Minimum Finding Algorithm [3] is the classical one and it is based on updating and broadcasting each smallest received value. A new leader selection algorithm for homogenous wireless sensor networks is presented in [4]. In [6], an improved version of the well-known Leader Election Ring algorithm is presented, where the authors to reduce the number of election messages by making assumptions on perfect clock synchronization and a perfect connection between transmitter and receiver. These assumptions are not realistic in wireless sensor networks and they require additional complex algorithms to deal with this synchronization. In [7], another improved version of Bully's algorithm is described, which acquires a smaller number of transmissions for leader election but takes more time. During leader election, a node will compare its value with the received value and only transmit the greater one. In [8], the network is divided and a pre-election to select a provisional leader is performed. The main drawback of this approach is when a node crashes and the contents of the memory will be completely removed. In [9], the number of nodes which can detect the failure is bounded before starting the Leader election algorithm. In this paper, the algorithm's time and complexity remain optimal even in worst case scenarios. The authors of [10] have modified Bully's algorithm in a way to improve the processing time. Their main contribution is that the election of a node is done on the basis of its performance and operation rate instead of on its higher identifier. Their algorithm allows to determine a leader before an existing leader dies. As in [6], the authors of [11] made the assumption of a perfect transmission connection and the time on air and collision scenarios is ignored, while implementing fault detection algorithms. Also, a series of dynamic leader election protocols in broadcast networks has been proposed. In [12], it is suggested to choose one leader and one leader assistant, so that in the case where the leader node crashes, the assistant node can take over the charge and coordinate other nodes. This can significantly decreasethe total number of elections

in the network, especially when the network size is large. A similar approach is followed in [5] whose approach does not rely on any particular network topology. In [13], two main algorithms are presented. The Bully algorithm [17] and the Ring Algorithm [14]. In [15], the author has proposed two algorithms working in the case of asynchronous networks. Both of the proposed algorithms can reduce the time complexity.

Except for the Minimum Finding Algorithm [3] and the algorithm of [15], the presented algorithms can be used only in the case of synchronous networks. In this article, we present a new algorithm that works with asynchronous networks without making any assumption on the topology. This algorithm has been compared with the classical Minimum Finding Algorithm, and the simulation results have shown that its complexity regarding exchanged messages is reduced by rates that can exceed 95%.

The remainder of the paper is organized as follows: In the following section, the Minimum Finding Algorithm will be reviewed. In Sect. 3, the Local Minimum Finding Algorithm (LMF) will be presented. Section 4 will present the proposed approach. The platform CupCarbon, which is used to implement the proposed algorithm, will be described in Sect. 5. In Sect. 6, simulation results will be presented. Finally, Sect. 7 concludes the paper.

2 The Minimum Finding Algorithm

In this section, we will present an algorithm that allows to determine a node leader representing the node with minimum or maximum value v. This value can represent the battery level, the residual energy, the identifier, the local energy, the x-coordinate in a network, etc. This algorithm is based on the Minimum Finding Algorithm presented in [3,16] which itself relies on the tree-based broadcast algorithm. The same algorithm can be used to find the maximum value. It can be described as follows. At the beginning, each node of the network assumes that its local value is the minimum of the network (the leader) and assigns it to the variable x_{min}. This value will be broadcasted and the corresponding node will wait for incoming x_{min} values from its neighbors. If a received value x_{min} is less than its local x_{min} value then this one will be updated and broadcasted again. This process is done repeatedly by each node as long as a received value is less than its local x_{min} value. After a given time t_{max}, only the leader, with the smallest value, will not receive a value that is smaller than its local x_{min} value.

Algorithm 1 is the pseudo-code of this process, where t_0 is the time of the first execution of the algorithm, that can correspond to the first powering-on of a sensor node, t_c the current local time of a sensor node, and t_{max} the maximally tolerated running time of the algorithm from the first execution to the current time of a sensor node.

Algorithm 1. *MinFind*: The pseudo-code of the classical Leader Election Algorithm

Input: t_{max}, v
Output: leader

```
1: leader = true;
2: t_0 = getCurrentTime();
3: x_min = v;
4: send(x_min, *);
5: repeat
6:     x = read();
7:     if (x < x_min) then
8:         leader = false;
9:         x_min = x;
10:        send(x_min, *);
11:    end if
12:    t_c = getCurrentTime();
13: until (t_c − t_0 > t_max)
```

In order to set the value of t_{max}, one needs to calculate the time complexity of this algorithm. For this purpose, let us consider the worst case represented by a linear network with n nodes, where we are searching for the node with minimum x-coordinate. This node, situated on the extreme left, will send only 1 message and will receive only 1 message. Nevertheless, the right-most node will receive and send $n-1$ messages of the received assumed x_{min} coordinate, since it is the node having the largest x-coordinate. Therefore, each one of the other nodes, except the extreme left one, has at least one node on its left. Thus, these nodes will broadcast the newly received x_{min}.

Altogether, the message complexity is equal to $M[MinFind] = 2(n-1) = 2n-2$. If we assume that a sensor node can send and receive messages simultaneously (full-duplex communication) the overall time complexity $T[MinFind] = n-1$. Since the time complexity is known, it is possible to estimate the value of t_{max}, the required time to find the leader. For example, in a network of 100 sensor nodes, with 1024 bits message size sampled with a 250 kb/s frequency (802.15.4 standard based network), 406 ms are required to find the leader. In this article, we have simulated two networks of 100 sensor nodes using the CupCarbon simulator. The first network is assumed to be linear (cf. Fig. 1) and second is assumed to be random (cf. Fig. 2). The simulation results show that the leader is captured in 406 ms with a consumption of 1 J to 9 J per node for the linear network. And in case of a random network, it took 70 ms with an energy consumption of 1 J to 5 J per node. In these simulations, the energy required for a serialization of data from the microcontroller to the RF radio module is neglected. But, if we assume a serialization time of 38400 b/s then to find the leader requires 1.5 s and 190 ms for the linear and the random network, respectively. Determining an accurate estimator of the value of t_{max} in the case of random networks could be a topic for future work.

Fig. 1. A linear network with 100 sensor nodes.

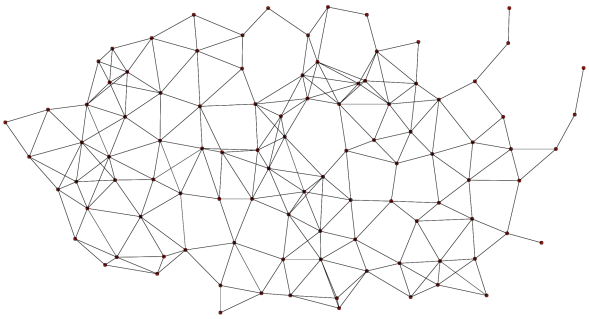


Fig. 2. A random network with 100 sensor nodes.

3 The Local Minima Finding Algorithm

A local minimum node, also called *Local Leader*, is the node which has no neighbor with a value smaller than its own value. But this value is not necessarily a global minimum. The marked nodes, represented by the red arrows in Fig. 4, show examples of local minima.

The Local Minima Finding (LMF) Algorithm uses the same principle as the previously presented *MinFind* algorithm to determine if a node is a local minimum or not, with the exception that each node will send its coordinates only once, and after receiving the messages from all its neighbors, it decides if it is a local minimum or not in case it receives a smaller value than its own. The algorithm of finding local minima is given as follows:

Algorithm 2. *LMF*: The pseudo-code of the Local Minima Finding Algorithm

Input: t, v
Output: local_min
1: local_min = true;
2: $x_{min} = v$;
3: send(x_{min}, *);
4: **while** (((x = read(t))$\neq$ *null*) and local_min) **do**
5: **if** ($x < x_{min}$) **then**
6: local_min = false;
7: **end if**
8: **end while**

4 The Proposed Method

4.1 Concept

In the *MinFind* algorithm each node is sending messages repeatedly and updates its values each time the received value is smaller than its own value. After a certain time, each node will be marked as a non-leader (or non-minimum) node, except the leader which has the smallest value since this node will never receive any smaller value than its own. This process is time-consuming and it requires a lot of broadcasting messages, which makes it very energy consuming and impractical in reality for the case of WSNs, because of collisions, for instance. To address this issue, we propose a new approach where each node will send a broadcast message once, in order to determine the local minima using the LMF algorithm (cf. Algorithm 2). Then each local minimum will send a message to a given reference node which will select the global minimum. This approach is detailed as follows:

Step 1: Mark each node as a global minimum and select one node as a reference node (cf. Fig. 3).
Step 2: Run the LMF algorithm to find the local minima nodes and unmark the other nodes (cf. Fig. 4).
Step 3: The reference node will send a message to the nodes in order to ask the local minima nodes to send their values (cf. Fig. 5).
Step 4: Each local minimum node will send a message to the reference node and the reference node will determine the global minimum from the received local minima nodes (cf. Fig. 6).
Step 5: The reference node will send a message to the global minimum node saying that it is the global minimum node (cf. Fig. 7).

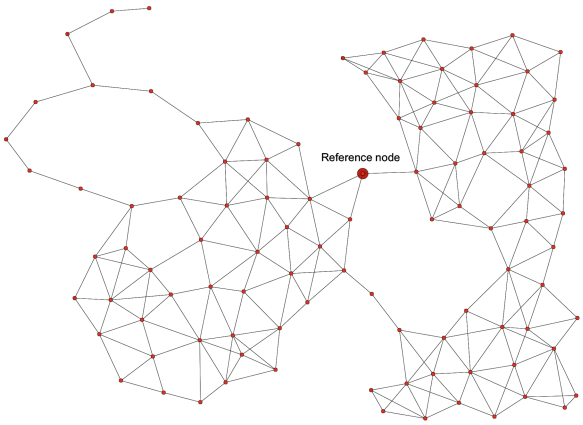


Fig. 3. Example of a network with a designed reference node.

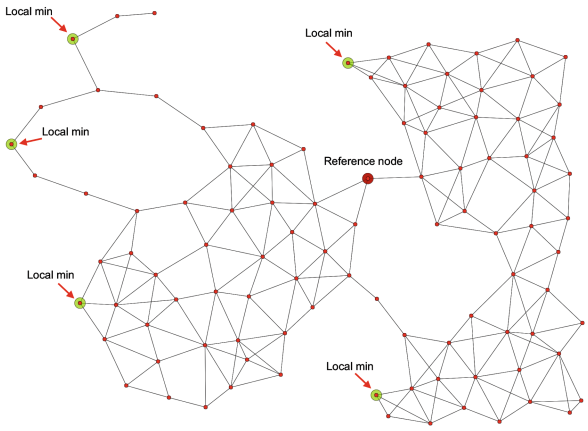


Fig. 4. The local minima found by Algorithm 2.

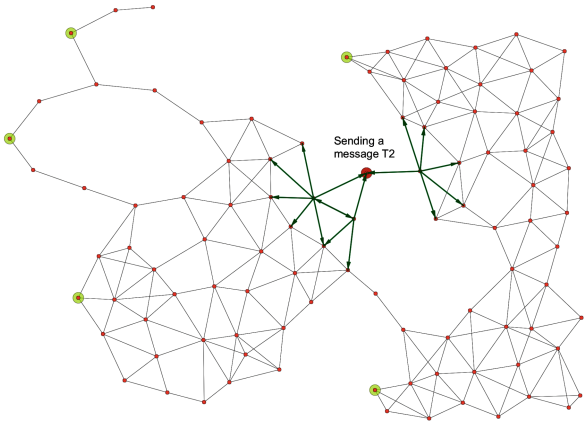


Fig. 5. The reference node asks for local minima nodes (flooding messages).

4.2 The LOGO Algorithms

To present the proposed algorithms, let us define in Table 1 some message primitives necessary for the communication between nodes and their definitions and in Table 2 the functions used in the algorithms. The proposed algorithm works in the case of bidirectional communication.

Note that there are two algorithms. One is executed by the reference node (Algorithm 3) and the second is executed by the remaining nodes (Algorithm 4). Algorithm 3 of the reference node takes as inputs a value x and the time wt required before selecting the global minimum. The output *global_min* is a variable which is equal to true if the current reference node is a leader (minimum) and false, otherwise. It starts with an initialization (lines 1 to 3). And it waits for 1 second (line 4), the necessary time to finish the process of determining the

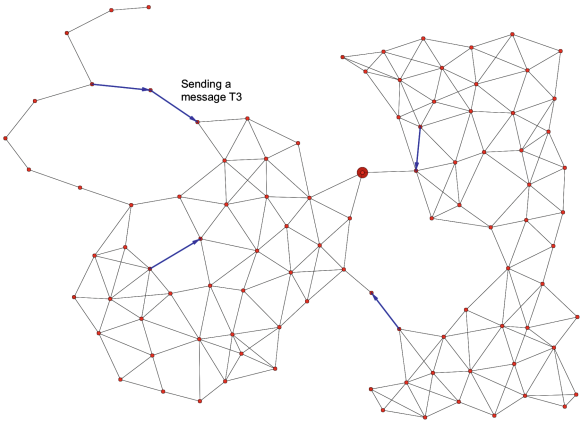


Fig. 6. The local minima nodes will declare themselves to the reference node (blue arrows). (Color figure online)

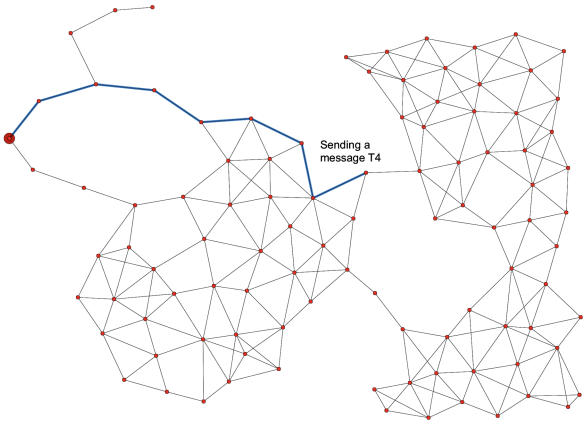


Fig. 7. The reference node designates the global minimum node (leader) and informs that node.

Table 1. Message primitives and their definitions.

Primitive	Definition
T1	I send you my value
T2	Send me your value if you are a local minimum
T3	I send you my value as a local minimum
T4	I want to inform the global minimum (leader election)

Table 2. Functions of the proposed algorithms.

Function	Definition
`getId()`	returns the node identifier
`delay(dt)`	waits *dt* milliseconds before going to the next instruction
`add(v,t)`	adds the value v at the top of the vector t
`pop(t)`	removes the value at the top of the vector t and returns it
`stop()`	stops the execution of the program
`send(a,b)`	sends the message a to the sensor node having the identifier b,
	or in a broadcast (if $b = *$)
`read()`	waiting for receipt of messages. This function is blocking, which means that if there is no received message any more, it remains blocked in this instruction
`read(wt)`	waiting for receipt of messages. If there is no received message after *wt* milliseconds then the execution will continue and go to the next instruction

local minima. This time must be changed if the number of neighbors of a sensor node is very important. It is the time required for any sensor node to send a message in a broadcast and receive mode from its neighbor nodes. Then it sends a message T2 to ask the local minima to send their value x (lines 5 and 6). In line 8, the reference node will wait for receiving a message containing the id of the transmitter (r_id), the value r_x of the local minimum message and t, the stack of the path from the local minimum node to the reference node. If a message is received before *wt* milliseconds, then it means that a message T3 is received from a local minimum node. In this case, the received value r_x is tested whether it is smaller than the current value x_min which at the beginning is equal to the local value x (line 18). If this is the case, the reference node will be declared as a non-global minimum (line 19), the value of id_min will be updated with the value of r_id (line 20), the value of x_min will be updated with the value of r_x (line 21) and the route t from the reference node to the local minimum (id_min) will be assigned to t_min (line 22). Otherwise, if the received message is null (line 9), which can happen when the node does not receive any message during the *wt* milliseconds, then the reference node has received messages from all the local minima. In this case, if the *global_min* value is equal to true, the reference node is the global minimum and the algorithm will stop (line 15). Otherwise, the route t is the one situated between the reference node and the global minimum node. A message T4 will be sent to the global minimum, having the identifier id_min, using the route t (lines 11 to 13) in order to elect it.

Algorithm 4 of the remaining node takes as input only the value x. The output *global_min* is a variable which is equal to true if the current node is a leader (global minimum) and false, otherwise. Each non-reference node starts with initializations (lines 1 to 5). The variable once1 is used to allow only once

Algorithm 3. *LOGO*: The pseudo-code of the reference node.

Input: x, wt
Output: *global_min*

```
 1: id = getId()
 2: x_min = x
 3: global_min = true
 4: delay(1000)
 5: message = (T2, id, null, null)
 6: send(message, *)
 7: while (true) do
 8:     (type, r_id, r_x, t)=read(wt)
 9:     if (type==null) then
10:         if (global_min==false) then
11:             n_id = pop(t_min)
12:             message = (T4, id_min, null, t_min)
13:             send(message, n_id)
14:         else
15:             stop()
16:         end if
17:     else
18:         if ((type==T3) and (r_x < x_min)) then
19:             global_min = false
20:             id_min = r_id
21:             x_min = r_x
22:             t_min = t
23:         end if
24:     end if
25: end while
```

the reception of T2 messages and the variable once2 is used to accept only once any received T4 message. Then it starts the process of the LMF by sending in a broadcast a T1 message in order to test if it is a local minimum or not by comparing the values received from its neighbors with its own value x. If any received value is smaller than its value, then the node will be considered as a non-global minimum (lines 9 to 15). Once all the values of the neighbors received, the algorithm goes to the second step, where it will wait for a message T2 initiated by the reference node. In this case, it will route this message to its neighbors and if it is a local minimum (*global_min* = true) then it will send the message T3 to answer the message T2 coming from the reference node, in order to tell him that it is a local minimum (lines 22 and 23). Finally, it will be considered as a non-global minimum (line 24). The next part of the algorithm concerns the creation of the route from the local minimum node to the reference node. If any node receives a message T3 then it will add itself to a stack t (line 28) representing the route from the local minimum to the reference node, and route it again to the node p_id which had sent him previously a T2 message (lines 29 and 30). As soon as all the non-reference nodes have done this step,

Algorithm 4. *LOGO*: The pseudo-code of the non-reference node.

Input: x
Output: *global_min*

```
 1: id = getId()
 2: x_min = x
 3: global_min = true
 4: once1 = false
 5: once2 = false
 6: message = (T1, id, x_min, null)
 7: send(message,*)
 8: while (true) do
 9:     (type, r_id, r_x, t) = read()
10:     if (type == T1) then
11:         if (r_x < x_min) then
12:             x_min = r_x
13:             global_min = false
14:         end if
15:     end if
16:     if ((type == T2) and (once1 == false)) then
17:         once1 = true
18:         p_id = r_id
19:         message = (T2, id, null, null)
20:         send(message, *)
21:         if (global_min == true) then
22:             add(id, t)
23:             message = (T3, r_id, x_min, t)
24:             send(message, p_id)
25:             global_min = false
26:         end if
27:     end if
28:     if (type == T3) then
29:         add(id, t)
30:         message = (T3, r_id, r_x, t)
31:         send(message, p_id)
32:     end if
33:     if ((type == T4) and (once2 == false)) then
34:         once2 = true
35:         if (r_id == id) then
36:             global_min = true
37:             stop()
38:         else
39:             n_id = pop(t)
40:             message = (T4, r_id, null, t)
41:             send(message, n_id)
42:         end if
43:     end if
44: end while
```

the reference node will be in the situation where he has received all the routes and values from the local minima and it chooses the one of the global minimum. Then it will send a message T4 to elect the global minimum (lines 11 to 13 of Algorithm 3). Finally, each non-reference node which receives a T4 message (line 32) containing the route t and the identifier r_id of the leader, will test if its identifier id matches the received identifier r_id (line 35). If yes, it will be elected (lines 36 and 37). Otherwise, it will route the same message to the next sensor node having the identifier n_id pulled from the route t (lines 39 to 41).

5 CupCarbon Simulator and SenScript

The simulation of networks is an essential tool for testing protocols and their prior performance deployment. Researchers often use network simulators to test and validate proposed protocols and algorithms before their real deployment. Indeed, such an establishment may be costly and challenging, especially when talking about a large number of nodes distributed at a large scale. This is why the simulation of networks is essential. CupCarbon is a Smart City and Internet of Things Wireless Sensor Network (SCI-WSN) simulator. Its objective is to design, visualize, debug and validate distributed algorithms for monitoring, tracking, collecting environmental data, etc., and to create environmental scenarios such as fires, gas, mobiles, and generally within educational and scientific projects. It can help to visually explain the basic concepts of sensor networks and how they work; it may also support scientists to test their wireless topologies, protocols, etc., cf. Fig. 8.

Networks can be designed and prototyped by an ergonomic and easy to use interface using the OpenStreetMap (OSM) framework to deploy sensors directly

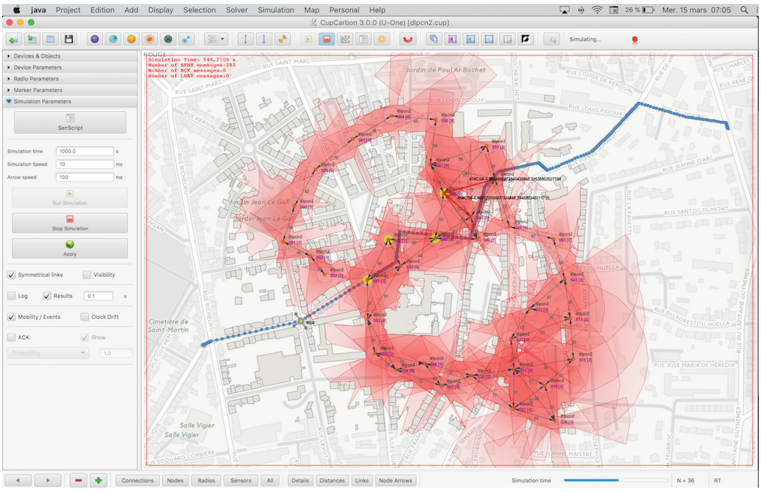


Fig. 8. CupCarbon user interface.

on the map. It includes a script called SenScript, which allows to program and configure each sensor node individually. The energy consumption can be calculated and displayed as a function of the simulated time. This allows to clarify the structure, feasibility and realistic implementation of a network before its real deployment. CupCarbon offers the possibility to simulate algorithms and scenarios in several steps. For example, there could be a step for determining the nodes of interest, followed by a step related to the nature of the communication between these nodes to perform a given task such as the detection of an event, and finally, a step describing the nature of the routing to the base station in case that an event is detected [18,19].

SenScript is the script used to program sensor nodes of the CupCarbon simulator. It is a script where variables are not declared, but can be initialized. For string variables, it is not necessary to use the quotes. A variable is used by its name, and its value is determined by $.

6 Simulation Results

In this section, we will compare the proposed algorithm with the classical *Min-Find* algorithm, since both of them can be used for any network. For the simulation, we have used the simulator CupCarbon [19], and SenScript is used to write the previously presented algorithms. We assume bidirectional communication between nodes. Figure 3 shows an example of a wireless sensor network designed in CupCarbon. We have randomly generated 10 networks with 20, 40, 60, 80, 100, 200, 300, 400, 500 and 600 sensor nodes, respectively. For each network, we have calculated the number of transmitted and received messages (exchanged messages) in order to compare their energy consumption which is directly related to this metric. We have obtained the graphs of Figs. 9 and 10. As we can see, the difference in each case can exceed 95% and this rate is increasing with the size

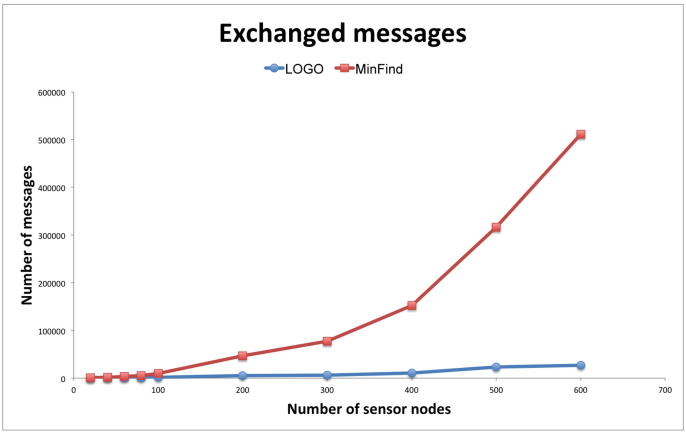


Fig. 9. Number of exchanged messages.

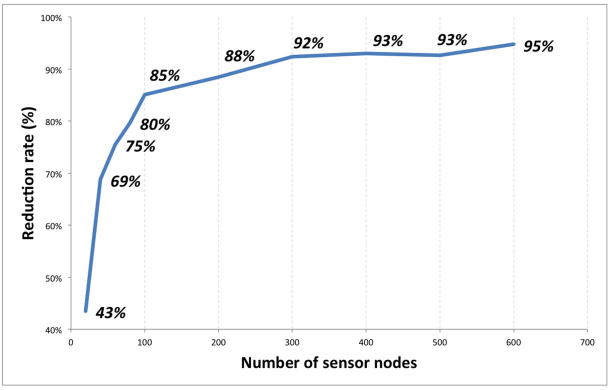


Fig. 10. Reduction rate in terms of the number of exchanged messages.

of the networks. From this figure, one can conclude that for very large networks, this reduction can reach 99%.

7 Conclusion

We have presented a new Leader Election algorithm which is low energy consuming. This algorithm is called LOGO (Local Optima to Global Optimum) where sensor nodes that are local leaders will send a message to a reference node which will designate the global leader and elect it by sending it a selection message. The classical algorithm allowing to find this node is called Minimum Finding Algorithm. In this algorithm, each node sends its value in a broadcast mode each time a better value is received. This process is very energy consuming and not reliable since it may be subject to an important number of collisions and lost messages. Our proposed algorithm is more reliable since broadcast messages are sent only twice by each node, and the other communications are based on a direct sending. The obtained results show that the proposed algorithm reduces the energy consumption with rates that can exceed 95% compared to the classical algorithm. We are now working on comparing our algorithm with other methods and to implement it on real hardware sensor platforms. Combining the proposed idea with some computational intelligence technique as presented in [20] can also give more favorable results in energy consumption.

Acknowledgment. This project is supported by the French Agence Nationale de la Recherche ANR PERSEPTEUR - REF: ANR-14-CE24-0017.

References

1. Saoudi, M., Lalem, F., Bounceur, A., Euler, R., Kechadi, M.T., Laouid, A., Madani, B., Sevaux, M.: D-LPCN: a distributed least polar-angle connected node algorithm for finding the boundary of a wireless sensor network. Ad Hoc Netw. J. **56**(1), 56–71 (2017). Elsevier
2. Lalem, F., Kacimi, R., Bounceur, A., Euler, R.: Boundary node failure detection in wireless sensor networks. In: IEEE International Symposium on Networks, Computers and Communications (ISNCC 2016), Hammamet, Tunisia, 11–13 May (2016)
3. Santoro, N.: Design and Analysis of Distributed Algorithms, vol. 56. Wiley, New York (2007)
4. Jabbar, S., Minhas, A.A., Gohar, M., Paul, A., Rho, S.: E-MCDA: extended-multilayer cluster designing algorithm for network lifetime improvement of homogenous wireless sensor networks. Int. J. Distrib. Sens. Netw., Article ID 902581, ISSN: 1550–1329 (Print), ISSN: 1550–1477 (Online)
5. Jabbar, S., Minhas, A.A., Imran, M., Khalid, S., Saleem, K.: Energy efficient strategy for throughput improvement in wireless sensor networks. Sensors **15**(2), 2473–2495 (2015). https://doi.org/10.3390/s150202473
6. Beaulah Soundarabai, P., Thriveni, J., Venugopal, K.R., Patnaik, L.M.: An improved leader election algorithm for distributed systems. Int. J. Next-Generat. Netw. **5**(1), 21 (2013)
7. Effat Parvar, M., Yazdani, N., Effat Parvar, M., Dadlani, A., Khonsari, A.: Improved algorithms for leader election in distributed systems. In: The 2nd IEEE International Conference on Computer Engineering and Technology (ICCET), vol. 2, pp. 2–6 (2010)
8. Kim, T.W., Kim, E.H., Kim, J.K., Kim, T.Y.: A leader election algorithm in a distributed computing system. In: Proceedings of the Fifth IEEE Computer Society Workshop on Future Trends of Distributed Computing Systems, pp. 481–485 (1995)
9. Chow, Y.C., Luo, K.C., Newman-Wolfe, R.: An optimal distributed algorithm for failure-driven leader election in bounded-degree networks. In: IEEE Proceedings of the Third Workshop on Future Trends of Distributed Computing Systems, pp. 136–141 (1992)
10. Park, S.H., Kim, Y., Hwang, J.S.: An efficient algorithm for leader-election in synchronous distributed systems. In: Proceedings of the IEEE Region 10 Conference TENCON, vol. 2, pp. 1091–1094 (1999)
11. Brunekreef, J., Katoen, J.P., Koymans, R., Mauw, S.: Design and analysis of dynamic leader election protocols in broadcast networks. Distrib. Comput. **9**(4), 157–171 (1996)
12. Zargarnataj, M.: New election algorithm based on assistant in distributed systems. In: IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2007, pp. 324–331 (2007)
13. Balhara, S., Khanna, K.: Leader election algorithms in distributed systems. J. Comput. Sci. Inf. Technol. IJCSMC **3**(6), 374–379 (2014)
14. Zargarnataj, M.: New election algorithm based on assistant in distributed systems. In: IEEE/ACS International Conference on Computer Systems and Applications, Amman, pp. 324–331 (2007)
15. Singh, G.: Efficient distributed algorithms for leader election in complete networks. In: 11th IEEE International Conference on Distributed Computing Systems, pp. 472–479 (1991)

16. Lynch, N.A.: Distributed Algorithms. Morgan Kaufmann, Massachusetts (1996)
17. Garcia-Molina, H.: Elections in a distributed computing system. IEEE Trans. Comput. **C–31**(1), 48–59 (1982)
18. CupCarbon simulator. http://www.cupcarbon.com
19. Mehdi, K., Lounis, M., Bounceur, A., Kechadi, T.: CupCarbon: a multi-agent and discrete event wireless sensor network design and simulation tool. In: IEEE 7th International Conference on Simulation Tools and Techniques (SIMUTools 2014), Lisbon, Portugal (2014)
20. Jabbar, S., Iram, R., Minhas, A.A., Shafi, I., Khalid, S., Ahmad, M.: Intelligent optimization of energy aware routing in wireless sensor network through bio-inspired computing: survey and future directions. Int. J. Distrib. Sens. Netw. **2013**, 13 (2013). Article Id 421084

An Efficient Privacy Preserving System Based on RST Attacks on Color Image

Sheshang D. Degadwala(✉) and Sanjay Gaur

Madhav University, Sirohi, Rajasthan, India
sheshang13@gmail.com, sanjay.since@gmail.com

Abstract. In Development of network communication need protect the transmission with fast Communication. Therefore, networking producers need to be constantly manage illegal use of the data. In our proposed approach, first step enter the user name and password then generate in text format that will be converted to QR-code using zxing library. Now the QR-code will be converted in to the share using Binary Visual cryptography algorithm. After that generated share-2 is save in the database that is for future reference at receiver side and share-1 is embedding into the R-Component LL bit using of block DWT-SVD and Pseudo Zernike moment. In embedded image further add G, B Component. So, Color watermark image is ready to transfer from the network. As in network there are different attackers apply RST attacks on the color watermark image and Generated attack Watermark Image. At the receiver side recover the attacks first apply Pseudo Zernike moment, Surf feature on R-component so, they will extract the attacks pixel and recover the scale-angle using affine transformation. Now share-1 and another share-2 is in data base so we will apply EX-OR operation to get the QR-Code. The final QR-code is decoded and we get the user name and password. This research work can give a way for providing authentication to all online Services.

Keywords: QR codes · VCS · RGB-extract · Block-DWT · Surf · Affine RST attacks

1 Introduction

In the world of communication, security assumes a basic part and claims a major management looking into its data. The announcement communication not withstanding a time's doesn't venture out alone, it will be attached unit with security parts. Subsequently security turns with make the way with open a correspondence box. Approaching data security, which is spread under cryptography, majority of the data hide or loss. Furthermore watermarking gives better part concerning with the handing sensitive data. Current Systems with the preventing methods continues evolving its face for boosted features, there may be need with get updated to it for its long run towards the improvement of future. Typically those happening is that the point when another calculation is transformed alternately an existing calculation is revised, intruders alternately hackers break the calculation. Along these lines it will be an absolute necessity on create

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 17–27, 2018.
https://doi.org/10.1007/978-3-319-73712-6_2

calculations that's only the tip of the iceberg proficient and make stable and unbreakable on the greater part degree. Normally, the organize security may be spread under cryptography and data concealing. Majority of the data concealing holds Steganography and watermarking which might be dated again old contrasted with cryptography. Done cryptography, scramble of data's takes put at those transmitter. Furthermore unscrambling them provides for those accurate enter toward those recipient area. Subsequently for scrambling Also unscrambling indicated actually Likewise encryption Furthermore decryption, a fact that utilized. Accordingly will scramble What's more unscramble same enter or distinctive keys might be utilized. Further extending its limbs under symmetric (conventional) Also deviated (public key) encryption. Here in the previous encryption, same fact that utilized both In those transmitter Also collector. Bit in the latter case, diverse fact that took care of. Advancing to Steganography, which will be craftsmanship of hiding of information under other. It might make dated past, yet once more heads should additional secure transmission. In place should enhance those security level, consolidation of Different systems will be took care of. One such attempt may be those Steganography again cryptography [2, 3].

We have made system to do secure transaction which is visual cryptography scheme and, for copyright protection and deal with geometrical attacks the watermarking scheme is used. It's absolutely impossible that anybody could decode the data contained inside some of shares. At the point when the shares are stack together, decoding is conceivable when the shares are set more than each other. Now, the data turns out to be in a flash accessible. No additional computational power is required keeping in mind the end goal to decode the data (Fig. 1).

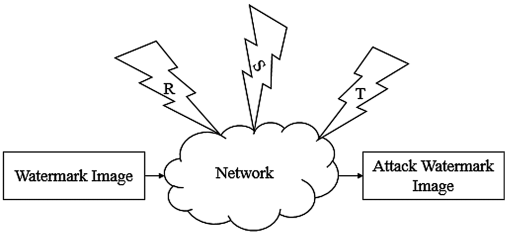


Fig. 1. Rotation (R), Scale (S) and Translation (T) attacks in network

Watermarking systems are arranged into spatial space techniques and change area strategies. Spatial area techniques are less unpredictable, however less strong against assaults. The watermarking plan in view of the change areas can be further Divided into discrete cosine transform (DCT), the discrete Fourier transform (DFT) and discrete wavelet transform (DWT). Capacity of DWT-SVD based plan is more than DFT.

A wide assortment of picture watermarking plans has been proposed and every locations a wide range of use situations.

2 Literature Survey

2.1 QR-Code [1, 4]

In computer networks development, distribution of "multimedia products is becoming gradually more day to day and the problems of digital copyright have become more and more famous. However, digital watermark is the new technology in the field of copyright protection. But it cannot effectively solve the problem of the arithmetical attacks in terms of image and the impact on the QR code fast responsive" characteristics [1].

Quick Response code is "2-dimension (2D) barcode, Denso Wave Corporation developed QR code in 1994. It can be improve the reading speed of 2D-barcodes and contains data for both vertical and horizontal dimensions and that's why it can contain a significantly greater amount of information. QD code contains information like text, web link, number, and multimedia data and is speed is 20 times faster than that of other 2D symbols. When secrete message embed into QR code, first it encode and then after develop the structure of QR code but it is time consuming, risky, and from QR code cannot get the secret message" directly [4].

2.2 VCS [2]

VCS is a new kind of cryptographic idea that efforts on resolving the problems of distribution the private images. VCS having the capacity to conceal information/data, for example, individual subtle elements is exceptionally fortunate. At the point when the information is covered up inside isolated pictures, it is altogether unrecognizable. At the point when the shares are partitioned, the information is totally ambiguous. Every picture holds distinctive bits of the information and when they are stacked together, the mystery message can be recuperated effortlessly. Every share relies on upon each other with a specific end goal to get the decoded data [2].

A pixel is a littlest component of an advanced picture. In a 32-bit advanced picture every pixel comprises of 32 bits, which is isolated into four sections, in particular red, green, blue and alpha; each with 8 bits. Alpha part introduces level of straightforwardness. In the event that each bits of Alpha part are '0', then the picture is absolutely straightforward. Human visual framework goes about as an OR work. In the event that two straightforward items are stacked together, then the last heap of articles will be straightforward. Be that as it may, in the event that one of them is non-straightforward, then the last pile of items will be non-transparent. Like 0 OR 0 = 0, considering 0 as straightforward and, 0 OR 1 = 1, 1 OR 0 = 1, 1 OR 1 = 1, in view of 1 as non-straightforward.

2.3 Digital Watermarking [5]

Digital watermark is "new approach to complement cryptographic processes. It is a visible or invisible identification code that is permanently embedded in the data and remains

present within the data after any decryption process [7]. The idea of computerized watermarking is gotten from steganography. Both steganography and watermarking plans are utilized to exchange data by implanting it into the" cover pictures [6].

A wide assortment of picture watermarking plans has been proposed and every locations a wide range of use situations. Watermarking systems are arranged into spatial space techniques and change area strategies. Spatial area techniques are less unpredictable, however less strong against assaults. The watermarking plan in view of the change areas can be further ordered into the Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) and so forth. Vigor is great in DWT based plan than DFT [5].

3 Proposed Preserving Method

After studying various visual cryptography schemes and watermarking schemes, we propose new technique for secure bank transaction. In this scheme we provide authenticity and data integrity of the shares using watermark technique. In our scheme we take one QR-image as original image or host image and create shares using 2-out-of-2 VC scheme [2]. When two shares will be created, server share is stored in bank database and client share is kept by user. The user will present with client share during all the transactions with bank. After that we apply the watermark technique on that client share image for providing the authentication and data integrity and send it on the open communication channel.

QR-Generation: As shown in the Fig. 2 first select the user name and password. Now using zxing library generating the QR-code. That QR-code is now in invisible form so now one can see the data inside. Further we have Apply VCS scheme to generate two shares of QR-Code.

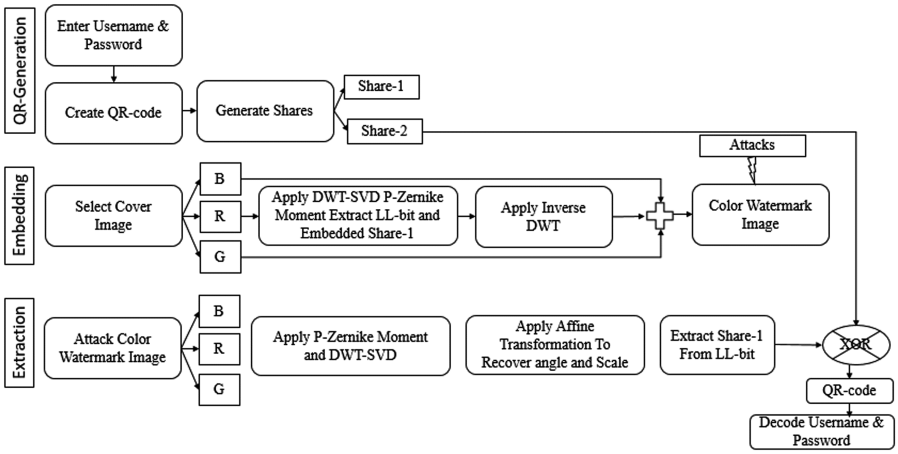


Fig. 2. Proposed flow

Embedding: In this process as shown in the Fig. 3 select the color cover image. Extract the R,G and B component. Now Select R-component and Apply P-Zernike Moment and DWT-SVD transformation and Extract LL-bit. In the LL-Bit embedding the Share-1 data. After Invers DWT-SVD transformation to generate R-Embedded Image Now Add Remain G and B Component to Create Color Water Mark Image. Color Watermark Image is transmitted over the Network Different Attackers Apply RST attacks on it.

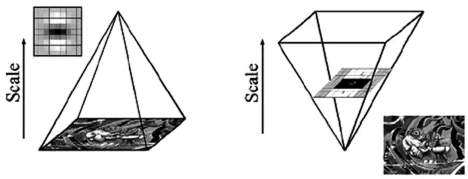


Fig. 3. Surf feature

Extraction: After RST attacks getting the Attack Color Image Which is now apply the P-Zernike Moment with Surf Feature Extraction to recover attacks. Now Extracting the share 1 and it will combine with another database share 2 to generate QR-image. QR decoder will decode the Username and Password.

The beauty of our system lies in the fact that, if any attacker makes a copy of any image share to forge it later, the watermark will be distorted so for such forged image share our system will not allow the generation of host image from the stack of 2 image shares. Thus, the attacker will not get the original image.

Here we use Singular Value Decomposition discrete wavelet transform based watermarking technique which is geometrically invariant. This type watermarking scheme is robust against the RST attacks, various JPEG and noise attacks.

3.1 Overall System

3.1.1 Encoder

Step 1: Enter User name and Password
Step 2: Encode to QR-Image
Step 3: Apply VCS and Generate 2-Share
Step 4: Share 2 is Save in Database
Step 5: Select Color Cover Image
Step 6: Extract R-Component
Step 7: Apply Block DWT + SVD + Pseudo Zernike Moment
Step 8: Embedding Share 1 in LL-band, G and B to Generate Watermark image

3.1.2 Network

Apply Rotation, Scale and Translation on Watermark Image.

3.1.3 Decoder

Step 1: Read Attack Watermark Image
Step 2: Extract R-Component
Step 3: Apply Pseudo Zernike Moment
Step 4: Apply Surf Feature Extraction and Affine Transformation
Step 5: Recover Rotation, Scale and translation Attacks
Step 6: Apply Block DWT + SVD
Step 7: Extract Share1 from LL-band
Step 8: Combine Share 1 and Share 2
Step 9: Decode QR-Image
Step 10: Recover User name and Password

3.2 VCS Algorithm

3.2.1 Share Generation

Generating those stakes from claiming mystery Image: in this stage usage from claiming Visual cryptography [2] may be completed. It includes those making of stakes starting with mystery picture utilizing (2, 2) VCS plan. Precise principal the mystery picture will be taken What's more is changed over should a double picture that point each pixel in the mystery picture is partitioned under eight sub pixels, four pixels in every impart Toward selecting the irregular pixel encoding plan crazy about scheme provided in algorithm.

3.2.2 Share Combination

In the keep going phase, those methodology for VCS mix may be performed. Here toward applying those double XOR operation, on both shares, we will get original data.

3.2.3 Embedding Algorithm

Step 1: Encode QR-image of Username and Password using Zxing 1.6 Library of java.
Step 2: Give Y a chance to signify the watermark inserting part, and utilize Haar orthogonal wavelet Transform to Y; then pick up the band LL which has most extreme vitality. Distribute LL into blocks Bi of size 4×4,

$$Z'' = [a_1, a_2, a_3 \ldots \ldots, a_s]$$

Where Zj'' is vector, and a_i is the SVD of all block, S is rank of all block.
Step 3: Apply the straightforward strategic monitor on encrypt the watermark.

$$x_{n+1} = \mu x_n(1 - x_n), 0 < x_n < 1, n = 0, 1, 2 \ldots ..10$$

Step 4: Calculate the value of Zj''

Norms $Zj'' = \sqrt{\sum_{j=1}^{s} a_j * a_j}$ and then $NO'' = \text{Norms } (Zj'')/D$.

Step 5: Embed bit using following technique.
If b = 1 then {if O is odd then O′ = O +1 else O′ = O} {Else {if E is even then E′ = E else E′ = E +1}}.

Step 6: Calculate the modified value and the modified vector as follows:

$$\text{Norms}\left(\text{Zj}'\right) = \text{NO}' \times \text{D} + (\text{D}/2),\ \text{Zj}' = \text{Zj} \times \text{Norms}\ (\text{Zj}')/\text{Norms}\ (\text{Zj})$$

Step 7: Apply inverse DWT to generate watermarked image.

3.2.4 Recover Decoding Algorithm

Step 1: To gauge the utilization of ensured Pseudo Zernike moments

$$S_{rc}(X, Y) = R_{rc}(X, Y)\exp\left(jm\tan^{-1}\left(\frac{X}{Y}\right)\right)$$

Where $X^2 + Y^2 \leq 1, r \geq 0, |c| \leq r$.

$$PZM_{rc} = \frac{r+1}{\pi}\sum X \sum Yf(X, Y)S_{rc}(X, Y)$$

A = absolute (Z)
Angle (Z) = tan − 1(imag(Z), real(Z));
Phi = angle (Z) * 180/pie

Step 2: Surf Feature exact [8]

Sense importance points, use Hessian matrix estimation. Form the integral pictures and the scale space of picture.

Importance point explanation and equivalent, descriptor defines the circulation of the intensity content, alike to SIFT. Based on sum of Haar wavelet reactions, construct a square region centered everywhere the interest point and concerned with along the location selected in earlier slice.

Step 3: pick up the Recovered watermarked image, and actualize 1-level DWT disintegration to its watermark embedding part. Get the sub-band LL′ which has incomparable vitality.

Step 4: Slice the sub-band LL″ into blocks Bi of size 4 × 4,

$$\text{Zj}'' = [\text{a}_1, \text{a}_2, \text{a}_3 \ldots \ldots, \text{a}_\text{s}]$$

Where Zj″ is a vector, and a_i is the SVD of all block and S is rank of all block.

Step 5: Calculate the value of Zj″,

Norms $\text{Zj}'' = \sqrt{\sum_{\text{j}=1}^{\text{s}} \text{a}_\text{j} * \text{a}_\text{j}}$ and the $\text{NO}'' = \text{Norms}(\text{Zj}'')/\text{D}$.

Step 6: Extract bit and extract watermark.
Step 7: Stacked extracted image with database share image with XORed operation.
Step 8: Decode QR-image to recover Username and Password.

4 Results and Discussion

As shown in Fig. 4 First user have to enter user name will enter by the user and Fig. 5 will be the QR-code generated by zxing library. Figure 6 is share 1 image generated by apply VCS algorithm.

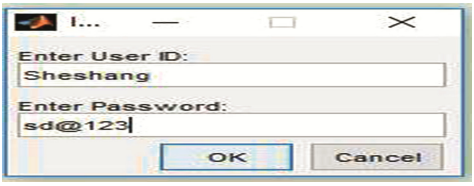


Fig. 4. ID & PSW

Fig. 5. QR-code

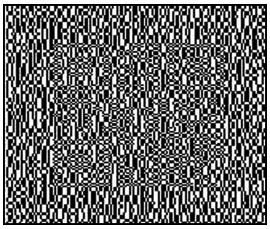


Fig. 6. Share 1

As shown in Fig. 7 Color image the DWT-SVD to getting the LL-bit as shown in Fig. 8. This image is now Combine with G and B to Create Color Watermark image. Attacker apply Rotation Attacks so getting the Fig. 9 image with rotation angle 30° (Table 1).

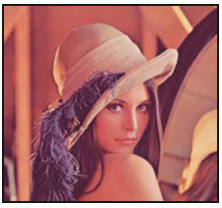


Fig. 7. Cover image

Fig. 8. DWT LL bit

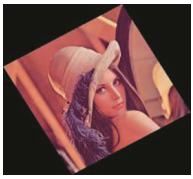


Fig. 9. Rotation 30°

Table 1. Results discuss

Rotation (Degree)	30°	45°	75°	90°	180°	270°
PSNR	63.21	63.13	63.03	Inf	Inf	Inf
MSE	0.026	0.026	0.025	0	0	0
Scaling	2	3	4	5		
PSNR	Inf	Inf	Inf	Inf		
MSE	0	0	0	0		
Translation	−10	−20	10	20		
PSNR	73.72	72.44	Inf	72.27		
MSE	0.0033	0.0055	0	0.0695		

Figure 10 shows the recover angle and Scale. Figure 11 will be generated by the P-pseudo Zernike and Surf Transformation. Figure 12 is Recover Share 1 from RST attacks. Figure 13 be the Recover the Username and password by decoding QR (Figs. 14 and 15).

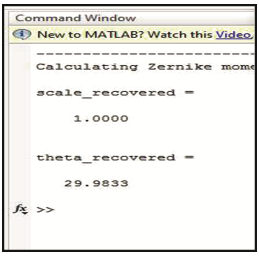


Fig. 10. Recover data

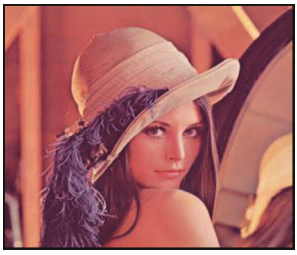


Fig. 11. Recover image

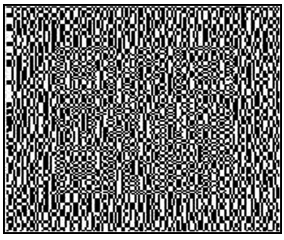


Fig. 12. Share 1

Fig. 13. Recover UID & password

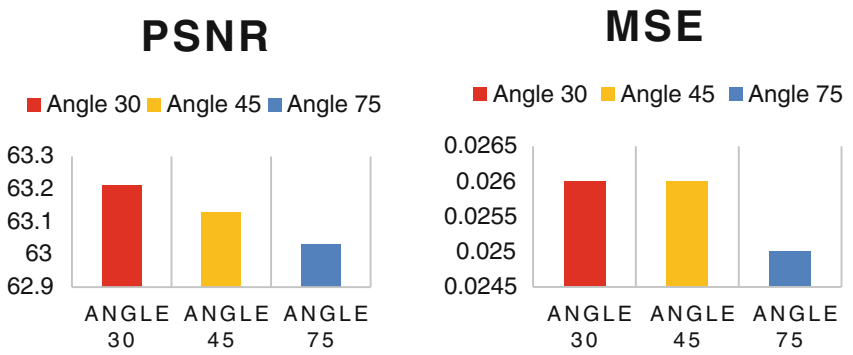


Fig. 14. Graphical representation of PSNR and MSE of Rotation Attacks

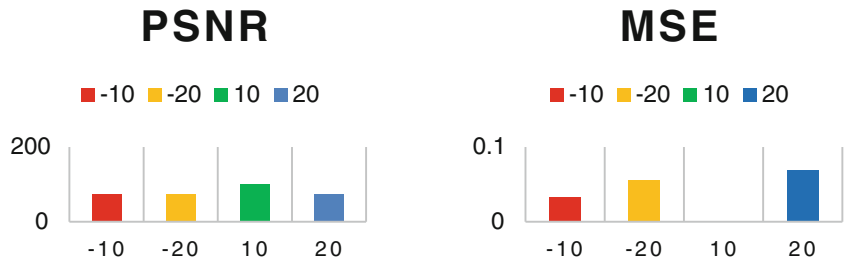


Fig. 15. Graphical representation of PSNR and MSE of translation attacks

5 Conclusion

In our proposed System we have convert username and password into QR-code. The QR-code is further divided into shares that shares are embedding into cover image. So its call multilayer Privacy. Now whenever Dual RST attacks apply on Color Cover image between transmission and receiving. Our Privacy Preserving System Recover Attacks. Here we have use Block DWT-SVD and Pseudo Zernike Moment with surf feature based watermarking system. Affine transformation is also apply for recover attacks on watermark image. So after extraction the proposed system will increase PSNR value for Recovered Image. This System Will Provide Efficient as well as Privacy Preserving Communication in Traditional Systems.

References

1. Delphin Raj, K.M., Victor, N.: Secure QR coding of images using the techniques of encoding and encryption. Int. J. Appl. Eng. Res. **9**(12), 2009–2017 (2014). ISSN 0973-4562
2. Ajish, S., Rajasree, R.: Secure mail using visual cryptography (SMVC). In: 5th ICCCNT 2014, 11–13 July 2014, Hefei, China (2014)
3. Gupta, A.K., Raval, M.S.: A robust and secure watermarking scheme based on singular values replacement. SaDhana **37**(4), 425–440 (2012)

4. Benoraira, A., Benmahammed, K., Boucenna, N.: Blind image watermarking technique based on differential embedding in DWT and DCT domains. EURASIP J. Adv. Sig. Process. **2015**, 55 (2015)
5. Gao, L., Gao, T., Sheng, G., Zhang, S.: Robust medical image watermarking scheme with rotation correction. In: Pan, J.-S., Snasel, V., Corchado, E.S., Abraham, A., Wang, S.-L. (eds.) Intelligent Data analysis and its Applications, Volume II. AISC, vol. 298, pp. 283–292. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07773-4_28
6. Nguyen, S.C., Ha, K.H., Nguyen, H.M.: An improved image watermarking scheme using selective curvelet scales. In: 2015 International Conference on Advanced Technologies for Communications (ATC) (2015)
7. Saxena, V.: Collusion attack resistant watermarking scheme for images using DCT. IEEE (2014)
8. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**, 346–359 (2007)

HiMod-Pert: Histogram Modification Based Perturbation Approach for Privacy Preserving Data Mining

Alpa Kavin Shah[1(✉)] and Ravi Gulati[2]

[1] MCA Department, Sarvajanik College of Engineering and Technology, Surat, India
alpa.shah@scet.ac.in
[2] Department of Computer Science, Veer Narmad South Gujarat University, Surat, India
rmgulati@vnsgu.ac.in

Abstract. Privacy Preserving Data Mining (PPDM) protects the disclosure of sensitive quasi-identifiers of dataset during mining by perturbing the data. This perturbed dataset is then used by trusted Third Party for effective derivation of association rules. Many PPDM algorithms destroy the original data to generate the mining results. It is essential that the perturbed data preserves the statistical inference of the sensitive attributes and minimize the information loss. Existing techniques based on Additive, Multiplicative and Geometric Transformations have minimal information loss, but suffer from reconstruction vulnerabilities. We propose Histogram Modification based method, viz. HiMod-Pert, for preserving the sensitive numeric attributes of perturbed dataset. Our method uses the difference in neighboring values to determine the perturbation factor. Experiments are performed to implement and test the applicability of the proposed technique. Evaluation using descriptive statistic metrics shows that the information loss is minimal.

Keywords: Privacy preserving data mining · Histogram Modification
Additive white Gaussian noise · Multiplicative perturbation
Geometric Data Perturbation

1 Introduction

Since last couple of decades, information collection over Internet is witnessing an exponential growth. More users have started providing their personal information in different Internet based activities like purchases/sales, auctions, entertainment, gaming, online surveys, to name a few. A person can now be easily and accurately linked based on his/her Internet activities, leading to a serious pose of privacy intrusion to the individuals. This vast pool of data has necessitated the need for efficient data mining protocols. Data mining which was limited and confined to narrower domain of Enterprises and Applications now encompasses Big Data and Cloud Computing.

Data collection has increased many-folds for research, trend analysis and more often collaborative mining results. It is vital that the information provided by the users should not breach their privacy. This concern has caught attention of researchers and is widely studied for improvements even today. PPDM algorithms tackle this issue by optimizing

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 28–36, 2018.
https://doi.org/10.1007/978-3-319-73712-6_3

privacy and minimizing information loss. This segment of Data Mining guard individual's privacy in Data Mining applications while providing accurate results for mining. An efficient PPDM algorithm must maximize privacy and minimize information loss. Also, it is desirable that the computational cost of generating perturbed data should be minimal and feasible. This requirement augments need of techniques that can be easily implemented by the contributing party. PPDM methods must protect the privacy of data and prevent adversaries to derive correlation between the distributed data. Our study focuses on developing a robust method which will be implemented by contributing party before releasing the data to Third Party. The proposed method will preserve the privacy of sensitive attributes and minimize the information loss.

1.1 Our Contributions

We make following contributions with this research:

1. We propose a Histogram Modification based method for perturbing numeric attributes for achieving privacy.
2. The resultant perturbed data obtained from proposed method is evaluated for efficiency in mining using statistical metrics. Also, a comparison of the proposed technique with basic perturbation techniques is conferred to show the effectiveness of our *HiMod-Pert* method.

1.2 Organization of the Paper

The rest of the paper is organized as follows:

- Section 2 insights the literature survey;
- Section 3 details the proposed *HiMod*-Pert method;
- In Sect. 4, we present the experimental results and present a comparison with contemporary perturbation techniques using descriptive statistical metrics;
- Finally, in Sect. 5, we provide conclusions and propose a road map for future work.

2 Literature Survey

Perturbation methods discussed by authors Domingo-Ferrer et al. [20] and Herranz et al. [21] are widely used in PPDM because the computational cost is lower than Cryptographic and Secure Multiparty Computations based methods. The former also has an edge over, as these methods can be used either by the data owner or by Trusted Third Party. Statistical Databases (SDBs) worked by authors Adam and Wortmann [1], Duncan and Mukherjee [2] and Gopal et al. [3] provide summary statistical information without sacrificing individual's sensitive identifying attributes. Numerical sensitive attributes of an application after perturbation must preserve the descriptive statistics for accurate mining. Perturbation Methods change the original data in a way that the summary statistics of the perturbed data remains same as that of the original data. For data mining to be effective, the perturbed data must preserve the relationships amongst

the contributing attributes. Authors Liu et al. [4] have exemplified the applicability of perturbation techniques by distorting the original values with known distribution, a category of probability distribution based perturbation approach. Authors Bai Li and Sarkar [7] have described a tree-based perturbation method. In this method, the original dataset values are replaced with fixed set of values. This technique is a type of fixed-perturbation based technique where values belonging to same group are replaced with certain defined values.

Perturbation methods Clifton et al. [5] and Kargupta et al. [6] broadly fall into three basic categories viz. Additive, Multiplicative and Rotation based methods like Geometric Data Perturbation (GDP). In Additive based method, first introduced by Agrawal and Srikant [10], randomized noise from known distribution sample like Gaussian is added to original data. If x_i is the original data values, and ε is random noise from some distribution like Gaussian or Uniform, new perturbed value $x_i + \varepsilon$ will appear instead of x_i. Many reconstruction approaches worked by authors Agrawal and Aggarwal [11], Domingo-Ferrer et al. [12] and Kargupta et al. [13] ascertain the vulnerability in privacy breaches with the use of Additive methods.

Another category of perturbation is Multiplicative based approach in which the Euclidean Distance is preserved well between the perturbed data and the original dataset. If x_i are the original data values and R is rotation matrix, perturbed values are computed as $R * x_i$. Independent Component Analysis (ICA) suggested by Liu et al. [15] when applied to perturbed values generated by multiplicative methods, can approximate original values. Work done by authors Liu et al. [14, 15] and Giannella et al. [16] suggest that the Multiplicative based methods have high privacy breach probability. Geometric based perturbations proposed by Chen et al. [17] add a random translation to values perturbed by Gaussian distribution. Their work enhances the resilience of random perturbation against three types of inference attacks: Naïve Inference attacks, ICA-based attacks and Distance-Inference attacks.

Our motive to present this study is to overcome the vulnerability due to randomized approaches and possible data reconstruction from original data. Researchers have given special attention to this and have presented novel studies for gaining knowledge from perturbed data. The recovery approach also is dependent on relative noise. The randomized approach of adding/multiplying Gaussian or Uniform noise to the original data sets does not ensure quality of data recovery process. In a cloud based environment it is essential at times to verify the integrity of the perturbed data. Sang et al. [22] have proposed and experimented reconstruction based on Undetermined Independent Component Analysis (UICA) where attacker has full or zero background information about perturbation matrix. Their studies clearly reveal the vulnerability of perturbation methods based on random and orthogonal projections. The authors' prior work Shah and Gulati [24] has revealed that the Additive and GDP based perturbation preserves the statistical inference of the original dataset and multiplicative perturbation methods generate records with minimum information loss but does not preserve statistical inference.

The Histogram Modification method suggested by authors Ni et al. [8] and Tai et al. [9] is a type of Data Hiding mechanism that works on images. It is a branch of Steganography where sensitive information is embedded into an image, making hiding imperceptible to humans. The image at receiver's end can be restored and the secret

information can be retrieved. The Histogram Modification technique is not suitable when images have equal Histograms. In Histogram Modification Technique data hiding is performed based on the difference of adjoining pixel values. To successfully retrieve the secret message and image, receiver must be passed the various peak points and zero points.

3 The Proposed Method

Rather than directly perturbing the values based on noise, we propose Histogram Modification based method for perturbing values. The proposed method does not add noise to all sensitive attributes like generic Additive and Multiplicative methods. Our proposed method uses difference in the adjoining neighbor values of dataset to generate noise which will then be added to or subtracted from the sensitive values. We have used peak as a measure of average of the difference of the adjoining data values. This peak value along with difference between adjoining data value is used to compute the perturbation factor. This perturbation factor will be different for each value. Unlike randomized Gaussian noise, this perturbation factor is dependent on the integrity of the dataset

Algorithm *HiMod-Pert*: Given the sensitive numeric attribute of a dataset, this algorithm returns its corresponding privacy preserving perturbed value.

Input: Let S, S = {s_i, i=1, 2, 3….n}, be the numeric sensitive attribute of Dataset D needed to be privatized before making D public.

Output: P is the perturbed sensitive attribute, P = {p_i, i=1, 2, 3…n} generated after applying ***HiMod-Pert*** method.

Step 1: [Calculate the difference between adjoining attributes of S. If the number of attributes n is odd, for the last attribute, the mean is used to calculate the difference.]

$$\text{diff}_i = \begin{cases} |s_i - s_{i+1}|, & i \neq n \\ |s_i - \frac{1}{n}\sum_{i=1}^{n} s_i|, & i = n \text{ and } n \text{ is odd} \end{cases} \quad (1)$$

Step 2: [Determine Peak from the difference obtained in Step 1]

$$\text{Peak} = \frac{1}{n}\sum_{i=1}^{n} diff_i \quad (2)$$

Step 3: [Calculate the Perturbation Factor p_{factor} for each of the sensitive attributes s_i.]

$$P_{factor_i} = (\text{Peak} - \text{diff}_i)/\text{Peak} \text{ for i } = 1,2,3 \ldots \text{n} \quad (3)$$

Step 4: [Apply the Perturbation Factor to s_i calculated in Step 3]

$$P_i = \begin{cases} s_i, & if\ i = 1\ or\ diff_i \leq Peak, \\ s_i + P_{factor_i}, & if\ diff_i > Peak\ and\ s_i \geq s_{i-1} \\ s_i - P_{factor_i}, & if\ diff_i > Peak\ and\ s_i < s_{i-1} \end{cases} \quad (4)$$

Fig. 1. Algorithm for proposed *HiMod-Pert* method

values. The first value of the dataset is not perturbed. In last step, the perturbation factor is added or subtracted to the original values based on the adjoining values.

Our proposed work does not embed message bit as our aim is to perturb the values and not hide any data. It can be extended to embed a message bit for increased privacy. The message bit must be shared between the contributing parties before perturbation. The integrity will be compromised with actions like deleting a sensitive record, changing the values, subsequently adding significant information loss to the mining results. Having briefed up the basic logic of the proposed method, we will now outline the algorithm of *HiMod-Pert* method based on Histogram Modification for applicability in privacy preserving data mining. We have considered that the sensitive attribute is application specific and can be identified using Decision Tree. The algorithm can be iteratively applied to perturb all the sensitive numeric attributes of the dataset. Figure 1 on subsequent page details our proposed HiMod-Pert algorithm.

4 Experimental Evaluation

4.1 Setting Environment

Experimental setup was done in MATLAB tool. The privacy attributes (columns) of the test data are determined by using Decision Tree suggested by authors Matatov et al. [25] and Fung et al. [26]. The Decision Tree sorts the columns by importance which can then be chosen for perturbation. Selection of four different datasets based on sizes of small, medium and large were chosen to test the performance of the proposed method. We have considered two datasets viz. ADULT and BREAST-CANCER–W from UCI Repository [18]. Both datasets contain large number of records and they exhibit real-world scenario. We have perturbed the numeric attributes Age and ID of the ADULT and BREAST CANCER-W dataset respectively. Another dataset HALD is available inbuilt with MATLAB. We have used INGREDIENTS dataset array from it as it is a Statistical Database. Lastly, we have also used NBASalaries dataset available from [19]. The attribute Salary was considered confidential. Table 1 describes the datasets used for our experimentation purpose and details number of instances and attributes. To provide a comparative analysis with basic perturbations, we have also simulated functions for Additive Perturbation, Multiplicative Perturbation and Geometric Data Perturbation in MATLAB. These methods are used as a baseline for comparison against our *HiMod-Pert* method.

Table 1. Datasets

Dataset	Number of instances	Number of attributes
ADULT	32561	15
BREAST-CANCER Wisconsin	699	10
INGREDIENTS	13	4
NBASalaries	407	6

4.2 Experimental Results

We have implemented the *HiMod-Pert* method in MATLAB. To show the performance of the proposed method, descriptive statistical measures like Mean, Standard Deviation, Mean Square Error, Root of Mean Square, Mean Absolute Error and Euclidean Distance are taken into consideration. For effective comparison, Table 2 consolidates the results generated on original Dataset, Additive, Multiplicative, GDP and our proposed *HiMod-Pert* method for various statistical metrics.

Table 2. Results of descriptive statistical measures on original and perturbed dataset generated by Additive, Multiplicative, GDP and *HiMod-Pert* method

Perturbation Techniques	Mean	Standard Deviation	Mean Square Error	Root of Mean Square	Mean Absolute Error	Euclidean Distance
ADULT DATASET						
Original DS	38.58	13.64	–	–	38.5816	–
Additive	38.59	13.67	0.98	41.00	38.60	178.87
Multiplicative	0.26	41.03	3.35e+03	41.16	31.00	1.04e+04
GDP	39.30	13.64	1.75	42.17	39.90	238.54
HiMod-Pert	***38.5449***	***13.3428***	***0.3634***	***40.7889***	***38.5449***	***108.7724***
BREAST CANCER-W DATASET						
Original DS	1.07e+06	6.17e+05	–	–	1.07e+06	–
Additive	1.07e+06	6.17e+05	0.99	1.23e+06	1.07e+06	26.30
Multiplicative	3.45e+04	1.05e+06	2.57e+12	1.18e+06	8.23e+05	4.25e+07
GDP	1.07e+06	6.17e+05	1.057	1.24e+06	1.07e+06	27.13
HiMod-Pert	***1.0352e+06***	***1.7688e+05***	***2.1669***	***1.0353e+06***	***1.0352e+06***	***7.2115***
INGREDIENTS DATASET						
Original DS	48.1538	15.5609	–	–	48.1538	–
Additive	48.4089	15.3281	0.5395	50.5994	48.4089	2.6484
Multiplicative	31.1067	27.6588	732.2388	40.9120	31.1067	114.9051
GDP	48.3421	15.7801	0.56	51.3433	48.0952	3.4452
HiMod-Pert	***48.0404***	***15.3437***	***0.2274***	***50.2514***	***48.0404***	***1.7192***
NBASALARIES DATASET						
Original DS	4.4695e+06	4.6933e+06	–	–	4.4695e+06	–
Additive	4.4695e+06	4.6933e+06	0.8221	6.4768e+06	4.4695e+06	18.2919
Multiplicative	3.6857e+06	5.5919e+06	4.8124e+13	6.6916e+06	3.6857e+06	8.3264e+07
GDP	4.5673e+06	4.6924e+06	0.3412	6.4523+06	6.4523+06	15.2347
HiMod-Pert	***4.4695e+06***	***4.6933e+06***	***0.6380***	***6.4768e+06***	***4.4695e+06***	**16.1143**

4.3 Experimental Inferences

Statistical Measures are used to check the applicability of perturbation techniques for information loss and privacy breach. The use of probabilistic information loss discussed by Mateo-Sanz et al. [23] is used to evaluate the information loss of the perturbed data. Mean, Standard Deviation (SD), Mean Square Error (MSE), Mean Absolute Error (MAE) and Root Mean Square (RMS) are used to evaluate the information loss for the

perturbed dataset. These statistical measures are necessary to prove the information loss for perturbed sets but not sufficient to conclude the same.

Mean and Standard Deviation measures the univariate information loss after perturbation. The experiments show that the Mean and Standard Deviation of original dataset is very near to perturbed dataset generated by our proposed *HiMod-Pert* method. This ensures that the proposed method efficiently preserves the clusters of original dataset.

Mean Square Error is a measure of average of squares of deviation of the original values from the perturbed values. For perturbed values to accurately estimate the original values, mean square error should be near to 1. The proposed *HiMod-Pert* method will generate MSE near to 1. Unlike Multiplicative method, it is efficient in preserving this statistical metric. Mean Absolute Error forecast how close are the perturbed values to the original values. It measures the distance between values generated by perturbation methods and original unperturbed values. The values in all the four datasets in our experimentation show that the Mean Absolute Error is same as that of original.

Euclidean Distance is a measure of how the values in perturbed dataset are linked with the original values. Smaller Euclidean Distance suggests that the probability of linkage of perturbed values to the original values is high. The proposed method uses adjacent values for finding the perturbation factor. Hence the record linkage is high. Both Additive and GDP methods have Euclidean Distance measures very less, indicating high record linkage. Root Mean Square of an estimator is the measure of imperfection of the fit of the perturbed data to the original data. For our *HiMod-Pert*, the value of Root Mean Square is effectively retained. The result, shown in Table 2 suggests that descriptive statistics is preserved well by *HiMod-Pert* method.

5 Conclusions

HiMod-Pert - a method based on Histogram Modification for effectively preserving the privacy and optimally minimizing information loss is proposed. We have exploited the traditional method that is used in Image Steganography. The method uses the differences in neighbouring sensitive attributes to modify the original values. Unlike contemporary methods where the transformation is fixed or based on randomization, we have suggested use of conditional perturbation factor that will be computed for each privacy sensitive attribute. Our experiments show that the method is effective for balancing between information loss and disclosure risk. Future work encompasses in studying the impact of various attacks, variations caused due to compromise in integrity and optimizing the method to combat against attacks.

References

1. Adam, N.R., Wortmann, J.C.: Security-control methods for statistical databases: a comparative study. ACM Comput. Surv. **21**(4), 515–556 (1989)
2. Duncan, G.T., Mukherjee, S.: Optimal disclosure limitation strategy in statistical databases: deterring tracker attacks through additive noise. J. Am. Stat. Assoc. **95**(451), 720–729 (2000)

3. Gopal, R., Garfinkel, R., Goes, P.: Confidentiality via camouflage: the CVC approach to disclosure limitation when answering queries to databases. Oper. Res. **50**(3), 501–516 (2002)
4. Liu, L., Kantarcioglu, M., Thuraisingham, B.: The applicability of the perturbation based privacy preserving data mining for real-world data. Data Knowl. Eng. **65**, 5–21 (2007)
5. Clifton, C., Kantarcioglu, M., Vaidya, J., Lin, X., Zhu, M.: Tools for privacy preserving distributed data mining. SIGKDD Explor. **4**(2), 38–44 (2002)
6. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: Random data perturbation techniques and privacy preserving data mining. In: IEEE International Conference on Data Mining (2003)
7. Bai Li, X., Sarkar, S.: A tree-based data perturbation approach for privacy-preserving data mining. IEEE Trans. Knowl. Data Eng. **18**(9), 1278–1283 (2006)
8. Ni, Z., Shi, Y.Q., Ansari, N., Su, W.: Reversible data hiding. In: Proceedings of International Symposium on Circuits and Systems, Bangkok, Thailand, vol. 2, pp. 912–915, 25–28 May 2003
9. Tai, W., Yeh, C., Chang, C.: Reversible data hiding based on histogram modification of pixel differences. IEEE Trans. Circ. Syst. Video Technol. **19**(6), 906–910 (2009)
10. Agrawal, R., Srikant, R.: Privacy preserving data mining. In: Proceedings of ACM SIGMOD Conference on Management of Data, Dallas, Texas, pp. 439–450, May 2000
11. Agrawal, D., Aggarwal, C.C.: On the design and quantification of privacy preserving data mining algorithms. In: Proceedings of the 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Santa Barbara, pp. 247–255 (2001)
12. Domingo-Ferrer, J., Sebé, F., Castellà-Roca, J.: On the security of noise addition for privacy in statistical databases. In: Domingo-Ferrer, J., Torra, V. (eds.) PSD 2004. LNCS, vol. 3050, pp. 149–161. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25955-8_12
13. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: Random-data perturbation techniques and privacy preserving data mining. Knowl. Inf. Syst. **7**(4), 387–414 (2005). https://doi.org/10.1007/s10115-004-0173-6
14. Liu, K., Giannella, C., Kargupta, H.: An attacker's view of distance preserving maps for privacy preserving data mining. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) PKDD 2006. LNCS (LNAI), vol. 4213, pp. 297–308. Springer, Heidelberg (2006). https://doi.org/10.1007/11871637_30
15. Liu, K., Kargupta, H., Ryan, J.: Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. IEEE Trans. Knowl. Data Eng. **18**(1), 92–106 (2006). https://doi.org/10.1109/TKDE.2006.14
16. Giannella, C., Liu, K., Kargupta, H.: Breaching Euclidean distance-preserving data perturbation using few known inputs. IEEE Trans. Knowl. Data Eng. **83**, 93–110 (2013). https://doi.org/10.1016/j.datak.2012.10.004
17. Chen, K., Sun, G., Liu, L.: Towards attack-resilient geometric data perturbation. In: Proceedings of the 2007 SIAM International Conference on Data Mining, Minneapolis, pp. 78–89 (2007)
18. Lichman, M.: UCI machine learning repository. School of Information and Computer Science, University of California, Irvine (2013). http://archive.ics.uci.edu/ml
19. https://github.com/Kjonge/DemoWorkbooks/blob/master/NBA%20salaries.xlsx
20. Domingo-Ferrer, J., Mateo-Sanz, J.M., Torra, V.: Comparing SDC methods for microdata on the basis of information loss and disclosure risk. In: Proceedings of the International Conference on New Techniques and Technologies for Statistics: Exchange of Technology and Knowhow, pp. 807–826 (2001)
21. Herranz, J., Matwin, S., Nin, J., Torra, V.: Classifying data from protected statistical datasets. Comput. Secur. **29**(8), 874–890 (2010). https://doi.org/10.1016/j.cose.2010.05.005

22. Sang, Y., Shen, H., Tian, H.: Effective reconstruction of data perturbed by random projections. IEEE Trans. Comput. **61**(1), 101–117 (2012)
23. Mateo-Sanz, J.M., Domingo-Ferrer, J., Sebé, F.: Probabilistic information loss measures in confidentiality protection of continuous microdata. Data Min. Knowl. Disc. **11**(2), 181–193 (2005). https://doi.org/10.1007/s10618-005-0011-9
24. Shah, A., Gulati, R: Evaluating applicability of perturbation techniques for privacy preserving data mining by descriptive statistics. In: Proceedings of 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, pp. 621–627, 21–24 September 2016
25. Matatov, N., Rokach, L., Maimon, O.: Privacy-preserving data mining: a feature set partitioning approach. Inf. Sci. **180**(14), 2696–2720 (2010). https://doi.org/10.1016/j.ins.2010.03.011
26. Fung, B.C.M., Wang, K., Yu, P.S.: Anonymizing classification data for privacy preservation. IEEE Trans. Knowl. Data Eng. **19**(5), 711–725 (2007)

Exhausting Autonomic Techniques for Meticulous Consumption of Resources at an IaaS Layer of Cloud Computing

Vivek Kumar Prasad(✉) and Madhuri Bhavsar

Nirma University, Ahmedabad 382481, Gujarat, India
{vivek.prasad,madhuri.bhavsar}@nirmauni.ac.in
http://www.nirmauni.ac.in/

Abstract. Internet based computing has provided lots of flexibility with respect to the usages of resources, as per the current demand of the users, and granting them the said resources has its own benefits, if given in proper manner i.e. exactly what the user has asked. In this paper the autonomic computing concepts has been discussed which will be very useful for the better utilisation of the resources at an IaaS (Infrastructure as a Service) level of the cloud computing. As Cloud Computing is highly scalable and virtualisation has become an important means for the efficient utilisation of the resources. Seeking the right amount of the resources at right time should be the goal of any CSP (Cloud service provider), On the other hand the CSPs has to deal with the situation of over provisioning and under provisioning, there should be some self-managing scheme through which the resources should be made available to the requesting user in an efficient manner to satisfy the need of their requirement with an improved resource utilisation. We have discussed the usage of autonomic computing to enhance the resource utilisation in the IaaS of cloud computing through various ways.

Keywords: Autonomic computing · Cloud computing
Over provisioning · Under provisioning · Resource utilisation
Infrastructure as a Service

1 Introduction

Cloud [22] is basic necessity today for the organisation because of its dynamic benefits and even though the organisation is 100% in cloud, it still requires the skilled professionals and managers, who can understand the right solutions where the cloud can be well suited to their enterprises. The reports indicates that the cloud will emerge in the future for betterment of the enterprise stakeholders, for handling the services [17] such as, IaaS (Infrastructure as a Service), PaaS (Platform as a service) and SaaS (software as a Service).

The resources at an IaaS level of the cloud the will give the illusion to the end user that, the an infinite pool of resources are present for the end user because of the term called as virtualisation.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 37–46, 2018.
https://doi.org/10.1007/978-3-319-73712-6_4

Every VMs should be associated to certain amount of resources available at an IaaS of cloud computing. Our aim should be to efficiently use these VMs so that there will be minimum wastage of the resources. In certain situations, their will be the wastage of resources [3].

The autonomic computing [12] can be classified into four parts i.e. self-configuration means the system should be adaptable with the changing environment, Self-healing which means discovering and diagnosing the problem in advance or heal the problem and prevent disruption, Self-Optimisation, i.e. tune the resources and workload to maximise the utilisation, Self-protecting i.e. to anticipate, detect and identify the problem and protect itself.

2 Objective

– Growing demand of the infrastructure has also increased the energy consumption. Our objectives will be to reduce the energy consumption using efficient resource usages. The resources has to be used in such a manner, so the even the last resource available at the data centre has to be used efficiently. To make use of the Autonomic Techniques to utilise the resources available at an IaaS level of the cloud computing.

3 Methodology

The Fig. 1 indicates the impact of autonomic computing at various levels/areas where, if the resources are used efficiently then, the energy will be reduced and the systems will have the properties of go green concepts and an increased revenue. The security, QoS Dynamic resource allocation, Iterative optimisation, Root cause analysis and energy efficiency can be combined with autonomic computing to efficiently use the resources available at IaaS.

3.1 Future Direction for Research in the Field of Resource Optimisation at the IaaS Level with the Help of Autonomous Computing

The autonomic computing [12], have four parts known as self-configuration, which means the system should be adaptable to the changing environment, Self-Healing, means the act to heal itself from the upcoming problem, self-optimisation indicates the handling of the system in such a way so that the resources should be maximally utilized, followed by self-protection, it handles the identification of security issues and protect itself.

Quality of Service: Cloud service providers (CSPs) need to ensure that sufficient amount of resources are available as per the requirement of the end users, the QoS [2] requirements Such as limit (deadline), response time (latency), and budget constraints should be encountered with the current requirement of the

Fig. 1. The taxonomy of the autonomic computing and efficient resource management in an IaaS level of cloud computing.

end user. These QoS foundation is derived from SLAs (Service Level Agreements), and if any violation happens w.r.t. the SLA, will lead to penalty and the QoS will be affected.

The DeSVi [8] architecture has been proposed for monitoring and detecting of any SLAs violation in cloud computing, the main components are VM deployer, which is responsible for allocation of tasks and its mapping to the available resources and another one is application deployer, which is responsible for the execution of the application and its metrics.

Security: Security [13] is an important aspects of cloud computing because of its distributed nature, as Differentiate between steady request verses DDOS attack, as if a coordinated attack is launched against the SaaS provider, the unforeseen rise in traffic might be erroneously assumed to be from the original/legitimate requests and resources would be scaled up to handle them. This in turn will increase running cost of application and also the wastage of the resources. Here again the resources (data storage and network hardware) will be misused.

Performance Anomalies [7]**:** In this paper, the concept of UBL (Unsupervised Behaviour Learning) has been discussed, which is a black-box unsupervised behaviour learning and is used to find out for the anomaly prediction of the system at IaaS clouds. The unsupervised Behaviour Learning influences the SOM (Self-Organizing Map) which captures the changing behaviour of cloud instances with human intervention. Based upon this learning techniques the UBL predict

previously unknown performance anomalies and delivers clues for any anomaly causes. This also target the scalable behaviour learning by making use of virtualizing and distributing the learning task to the distributed hosts.

Energy Efficiency [4]**:** As cloud will give the illusion to the end user that cloud provides infinite pool of resources and on the other side the cloud needs to look for efficient usages of the energy also, example include as over provisioning of the resources, minimising the carbon footprint and server consolidation. Applications need to be scheduled in such a way, so that their total energy consumption is minimized by the effective usages of the resources at IaaS level. There are various research has been done on this field, some of them has been discussed below:-

In paper [9] considers the SLA with effective VM placement is used to minimize the operational cost in the cloud computing system which in turn reduces the energy consumption by considering the general queuing theory models for real world workloads.

In this research paper [5] the energy has been reduced and there is also a decrease in resource wastage by fixing up the predefined SLAs (Service Level Agreements) i.e. the SLA violation = 0% and the mechanism that has been used here are as Monte Carlo and random methods.

Here the authors [6] has mentions about the advantages of using a good architecture for energy efficient cloud computing and also discussed about the algorithms that are energy aware and should work within the limits of the SLAs.

Dynamic Resource Allocation: Scaling in/Scaling out [28] (that is the expanding/shrinking of resources, also called as elasticity and is one of the important property of the cloud computing), has to be carried out w.r.t. changing demands of the end users. The dimensions of the resources that has to be considered are, number of CPUs, amount of memory and size of the virtual disk at IaaS level of cloud computing. In the research paper [14], the CometCloud autonomic cloud engine concept has been highlighted that works on policy based mechanism, The keywords Autonomic cloudbridging and cloudbursting has been described, the cloudbridging merges the computation of local environment (i.e. grids and datacenters) and public cloud services (i.e. Eucalyptus and Amazone EC2) as well, the autonomic cloudbursting allows spikes in demand and dynamic application scale out for dynamic workloads.

Prediction for Resource Selection: The total cost of the resources depends on the type of the resources made available to the end user or resources provisioned, a prediction mechanism should be realised, and that takes into account of historical data execution statistics, for the fulfilment of the resources demand. If the predicted resources estimates are correct, then the wastage of the resources are minimised.

In this paper [25] authors have discussed about the workload forecasting and about the optimal resources allocation, challenged involved in autoscaling, predictive algorithms for autoscaling and its empirical results which satisfies the QoS and less operational cost.

The ecosystem of existing Big data tools [11], the analytics today require the support for big data and its implementation in the cloud computing, As there are various open source technologies related to big data analytics are available such as Spark (analytics), Hive and Pig, Storm (used in stream processing), YARM (MapReduce and other parallel programming)/Hadoop, HDFS (File systems and NOSQL databases), Cassandra and CouchDB.

The main aim is to identify how to use these tools, so that minimum resources has to be used (without the wastage of the resources) and no SLAs violation.

In [1] authors have discussed the state of the art of scalable data management for cloud computing infrastructure for heavy and analytical workloads. The designing of the data management has been highlighted. Apart from this different multitenancy models in the database has also been identified.

Scalable Decision Making Algorithms: Will be used for Big data analytics kind of scenarios [20] and it has to be recognised using scalable data management systems which uses machine learning, artificial intelligence, decision making and data mining techniques in clouds.

Here in this research paper the authors [29] have mentioned about the migration scenarios in the cloud, as migrations can be done in seconds or sometime more than this depending on the size, work type and bandwidth of the VMs and physical machines, the migration techniques should be automated through application environment with pre-defined strategy, and less human intervention. The two steps for implementing autonomous concepts are to adopt a distributed architecture where resource management is decomposed and each tasks is performed by Autonomous Node Agents, which are tightly coupled with the physical machines.

Root Cause Analysis/Identifying the Co-relation: In the real application scenario of cloud computing, the changes done at one end can affect the other end too, because of coupling hence mining dependency between anomalies of two different application layers are an important promising research direction. If the dependency has been identified, then it can be modelled maintained in the form of knowledge representation languages in the system premises and is known as knowledge-base.

The paper [21] discuss about the formal mathematical based decision model which established a logical chain of services requirements, the basic aim is to determine which cloud provider is well suited w.r.t. the requirement of the users, based upon these analysis, the risk are modelled by keeping view of integrity, availability and confidentiality.

Multi-resource Anomaly Detection: If multiple resources are identified for an anomaly detection then its an advantage, as its nice to find out the target resources which contributes more to the anomalies, that has been detected in the application and that affects the QoS, on the other side considering only one resource at a time causes needless delay.

In this paper [24] the Intrusion Detection and Prevention Systems (IDPS) and alarm management techniques has been discussed, the hardware and software

changes constructs the autonomic managers monitoring scheme and it optimise its use of resources without any human intervention.

Software defined network have current challenges as efficient utilisation of storage and network- resources utilisation at IaaS level through mathematical and statistical methods.

Here the authors [23] have discussed about the autonomous software defined network which make use of agent based architecture which controls the network, and the network controller generates events every time, when the network changes its state, and these events can be used to update the states or the facts in the knowledge base.

Failure Management in cloud computing, which is also known as Fault tolerance for performance optimisation or grace full degradation which in turn results as robustness, stability and reliability of the system, lead to the better utilisation of the resources.

The authors [15] have proposed an efficient fault management mechanism, which is based upon cognitive control loops, which uses alarm dataset, which consists of the data that has been inserted at the time of learning phase of the control loops, the SWRL (Semantic Web Rule Language) and Ontology are used to show the relationship between the alarm and its consequent services/actions, even the alarms are used w.r.t. there priority and is useful for finding the root causes easily by imposing the concept of associated rule miming techniques.

The Iterative Optimisation, which entitles the subset of analytic task, the functions which is being repeated, need to be profiled, information should be reserved, so that the cost and execution time will be known in advance and the resources can be utilised properly.

In this paper [10] the authors have proposed the task scheduling optimisation algorithms, with minimised of cost of resources used. They have used the technique of PSO (Particle swarm optimisation) with crossover, mutation and local search algorithms.

The paper [18] discussed about the QoS and its optimisation s design requirements, the application software should be adaptable to different run time situations, the cloud should also have infrastructure for deploying monitoring applications based on the QoS parameters and there should be a cloud feedback loop, that must be able to track the performance model and also to be used for management decisions.

In paper [16] the scheduling of dataflow operations has been considered, as minimum completion time in a given budget, than to minimise the monetary cost given a deadline and the differences between completion time and monetary cost, these values will be used in approximate optimisation framework which deals with the elasticity in the cloud. The practice proposed in the final step of the optimization of the system and its all of the parameters are instantiated automatically using functions or statistics collected during previous executions. System agents [19] (embedded in to system), monitoring CPU, schedulers, optimisation of the dynamic resource handling techniques, here the different agents will interact with each other for Self-Management.

In paper [27] the Multi Agents System concepts has been used, where agents will interact with each other for intelligent behaviour which result into a flexible cloud, autonomic and scalable and it also mentions about the techniques and methodology that account for the changing states and its behaviour. In an another paper [26], the concept of multi agent system has been included in the cloud environment for achieving the goal of implementing high performance complex systems and intelligent applications, the intelligence can be achieved by incorporating dynamic, flexible and autonomous behaviour.

The Generic Architecture and Its Approach: The autonomic computing techniques has been used to automate the operations in the cloud with minimal intervention of human interaction. As if multi resources anomaly detection can be handled using autonomic techniques then the resources can be used with its maximum utility, same is the case with SLA management with respect to the QoS parameters and others, same has been discussed above. The correct prediction will enhance our system and will lead to the optimum usages of the resources.

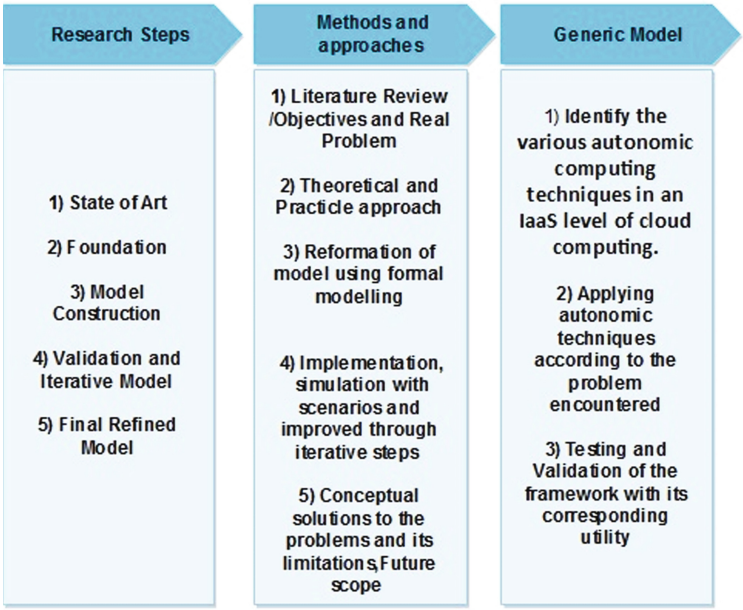


Fig. 2. Research approach towards the generic model.

The Fig. 2 discuss about the approach to the generic model, which initially starts from the state of art and move along the various methods and approaches which finally leads to the generic model, that should consists of the following features:-

– The concepts such as QoS metrics, unsupervised behaviour, queuing theory, Monte Carlo methods can be used for energy efficient resource management. The ideas of policy based management, dynamic resource demand, machine learning, artificial intelligence, data mining techniques decision making using pre-defined strategy and autonomous node agents can be used for the process of predicting resource selection and scalable decision making algorithms. Alarm management techniques, SDN agents and raising alarm with its related action decision approaches with monitoring mechanism can add benefits to the automatic anomaly detection mechanism.

Similarly new avenues has to be explored for making the cloud fully autonomous. The pairing of the autonomous techniques to its branch technology has to be analysed and implemented to make worth utilisation of the resources available at an IaaS level.

4 Conclusion

The autonomic computing mechanism to cloud computing will be a boon to the current technological scenario where cloud computing concepts are being used, as this will reduce the human intervention and the system itself will take wise decisions based upon the dynamism but still we have lots of open issues and challenges. In this paper we have highlighted a number of questions which can be used as a research agenda for making full utilisation of the resources available at an IaaS level of cloud computing by exhausting autonomic techniques. We have also identified the solution for the above problem and also mentioned some basic solutions which can, if implemented properly will result into the better utilisation of the resources.

References

1. Agrawal, D., Das, S., El Abbadi, A.: Big data and cloud computing: new wine or just new bottles? Proc. VLDB Endow. **3**(1–2), 1647–1648 (2010)
2. Al-Shehri, S.F.S., Li, C.L.: Quality of service for cloud computing. In: Advanced Materials Research, vol. 905, pp. 683–686. Trans Tech Publications (2014)
3. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., et al.: A view of cloud computing. Commun. ACM **53**(4), 50–58 (2010)
4. Berl, A., Gelenbe, E., Di Girolamo, M., Giuliani, G., De Meer, H., Dang, M.Q., Pentikousis, K.: Energy-efficient cloud computing. Comput. J. **53**(7), 1045–1051 (2010)
5. Borgetto, D., Maurer, M., Da-Costa, G., Pierson, J.-M., Brandic, I.: Energy-efficient and SLA-aware management of IaaS clouds. In: Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet, p. 25 (2012)
6. Buyya, R., Beloglazov, A., Abawajy, J.: Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. arXiv preprint arXiv:1006.0308 (2010)

7. Dean, D.J., Nguyen, H., Gu, X.: UBL: unsupervised behavior learning for predicting performance anomalies in virtualized cloud systems. In: Proceedings of the 9th International Conference on Autonomic Computing, pp. 191–200. ACM (2012)
8. Emeakaroha, V.C., Netto, M.A.S., Calheiros, R.N., Brandic, I., Buyya, R., De Rose, C.A.F.: Towards autonomic detection of SLA violations in cloud infrastructures. Future Gener. Comput. Syst. **28**(7), 1017–1029 (2012)
9. Goudarzi, H., Ghasemazar, M., Pedram, M.: SLA-based optimization of power and migration cost in cloud computing. In: 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp. 172–179. IEEE (2012)
10. Guo, L., Zhao, S., Shen, S., Jiang, C.: Task scheduling optimization in cloud computing based on heuristic algorithm. JNW **7**(3), 547–553 (2012)
11. Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U.: The rise of big data on cloud computing: review and open research issues. Inf. Syst. **47**, 98–115 (2015)
12. Huebscher, M.C., McCann, J.A.: A survey of autonomic computing degrees, models, and applications. ACM Comput. Surv. (CSUR) **40**(3), 7 (2008)
13. Hwang, K., Li, D.: Trusted cloud computing with secure resources and data coloring. IEEE Internet Comput. **14**(5), 14–22 (2010)
14. Kim, H., Parashar, M.: Cometcloud: an autonomic cloud engine. In: Cloud Computing: Principles and Paradigms, pp. 275–297 (2011)
15. Kim, S.-S., Seo, S.-S., Kang, J.-M., Hong, J.W.-K.: Autonomic fault management based on cognitive control loops. In: 2012 IEEE Network Operations and Management Symposium (NOMS), pp. 1104–1110. IEEE (2012)
16. Kllapi, H., Sitaridi, E., Tsangaris, M.M., Ioannidis, Y.: Schedule optimization for data processing flows on the cloud. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, pp. 289–300. ACM (2011)
17. Kundu, A., Banerjee, A., Saha, P.: Introducing new services in cloud computing environment. In: International Journal of Digital Content Technology and Its Applications, AICIT. Citeseer (2010)
18. Li, J., Chinneck, J., Woodside, M., Litoiu, M., Iszlai, G.: Performance model driven QoS guarantees and optimization in clouds. In: Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, pp. 15–22. IEEE Computer Society (2009)
19. Lopez-Rodriguez, I., Hernandez-Tejera, M.: Software agents as cloud computing services. In: Demazeau, Y., Pěchouček, M., Corchado, J.M., Pérez, J.B. (eds.) Advances on Practical Applications of Agents and Multiagent Systems. Advances in Intelligent and Soft Computing, vol. 88, pp. 271–276. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19875-5_35
20. Low, C., Chen, Y., Mingchang, W.: Understanding the determinants of cloud computing adoption. Ind. Manag. Data Syst. **111**(7), 1006–1023 (2011)
21. Martens, B., Teuteberg, F.: Decision-making in cloud computing environments: a cost and risk based approach. Inf. Syst. Front. **14**(4), 871–893 (2012)
22. Mell, P., Grance, T., et al.: The NIST definition of cloud computing (2011)
23. Passito, A., Mota, E., Bennesby, R., Fonseca, P.: AgNOS: a framework for autonomous control of software-defined networks. In: 2014 IEEE 28th International Conference on Advanced Information Networking and Applications (AINA), pp. 405–412. IEEE (2014)
24. Patel, A., Taghavi, M., Bakhtiyari, K., JúNior, J.C.: An intrusion detection and prevention system in cloud computing: a systematic review. J. Netw. Comput. Appl. **36**(1), 25–41 (2013)

25. Roy, N., Dubey, A., Gokhale, A.: Efficient autoscaling in the cloud using predictive models for workload forecasting. In: 2011 IEEE International Conference on Cloud Computing (CLOUD), pp. 500–507. IEEE (2011)
26. Talia, D.: Cloud computing and software agents: towards cloud intelligent services. In: WOA, vol. 11, pp. 2–6. Citeseer (2011)
27. Talia, D.: Clouds meet agents: toward intelligent cloud services. IEEE Internet Comput. **16**(2), 78–81 (2012)
28. Xiao, Z., Song, W., Chen, Q.: Dynamic resource allocation using virtual machines for cloud computing environment. IEEE Trans. Parallel Distrib. Syst. **24**(6), 1107–1117 (2013)
29. Yazir, Y.O., Matthews, C., Farahbod, R., Neville, S., Guitouni, A., Ganti, S., Coady, Y.: Dynamic resource allocation in computing clouds using distributed multiple criteria decision analysis. In: 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD), pp. 91–98. IEEE (2010)

Efficient Resource Monitoring and Prediction Techniques in an IaaS Level of Cloud Computing: Survey

Vivek Kumar Prasad(✉) and Madhuri Bhavsar

Nirma University, Ahmedabad 382481, Gujarat, India
{vivek.prasad,madhuri.bhavsar}@nirmauni.ac.in
http://www.nirmauni.ac.in/

Abstract. In this paper, we have discussed about the various techniques through which the cloud computing monitoring and prediction can be achieved, This paper provides the survey of the techniques related to monitoring and prediction for the efficient usages of the resources available at the IaaS level of cloud. As cloud provides the services, which are elastic, scalable or highly dynamic in nature, which binds us to make the correct usages of the resources, but in real situations the (Cloud Service Provider)CSP's has to face the situation of under provisioning and over provisioning, where the resources are not fully utilized and being wasted, though this is the survey paper, it ends up with the proposed model where both the concepts of the Monitoring and Prediction will be combined together to give a better vision of the future resource demand in IaaS layer of Cloud Computing.

Keywords: Cloud computing · Monitoring · Prediction
Under provisioning · Over provisioning · IaaS

1 Introduction

Cloud computing [24] is a techniques which allows suitable, on demand network access to the pool of various computing resources such as network, servers, storage application and other services, that can be quickly given back to the end user and released with minimal management effort. The cloud computing services [30] can be classified as Software as a Service Platform as a Service and Infrastructure as a Service along with different deployment models [8]. Essential characteristics of Cloud Computing [14,27] are On-Demand Self-Service, Broad Network Access, Resource pooling, Rapid elasticity, Measured service and metering and billing.

The resource management has to be efficiently used in the IaaS level of cloud computing, because the resources has to be allocated in a right amount [36] for an application. The interconnected resource management areas for efficient resource management are [5], resource discovery, Resource modeling, resource

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 47–55, 2018.
https://doi.org/10.1007/978-3-319-73712-6_5

Table 1. The Literature survey on Monitoring in to the cloud computing for efficient resource utilization

Sr.No	Authors	Objectives and methodologies	Conclusion and future directives
1	Whiteaker et al. [35]	To identified the delay measurements of the virtual machines (VMs) that consume CPU, memory, I/O, hard disk, and network bandwidth. As Heavy network usage of these competing VMs can introduce high round-trip times	On the spot decision to select the appropriate scheduling algorithms based upon the current scenario has to be identified. Dynamic Scheduling algorithms has to be used to deal with these kinds of issues
2	Wang [34]	A system integrating monitoring with analytics, termed as "Monalytics"? has been discussed which can capture, aggregate, and incrementally analyse data on demand. The properties of the Monalytics are as follow:- 1. Zooming in to 'interesting' locations at regular periods of time. 2. Reducing 'Time to Insight' i.e. capturing the total delay between when 'interesting' events occur and by the time they are recognized (i.e., after analysis is complete)	Identifying patterns usage and. finding ways to reduce Datacentre energy use. Fault Patterns or cost /effectiveness needs
3	Clayman et al. [11]	The distributed model has been used which consists of Virtualisation Plane, the Management Plane, the Knowledge Plane, the Service Enablers Plane, and the Orchestration Plane. Working together these distributed systems form a software-driven virtual network control infrastructure that runs on top of all current network and service infrastructures	Monitoring should not affect the performance and account in case of elasticity, scalability, federation and adaptability without violating the performance instances
4	Hasselmeyer and d'Heureuse [16]	The architecture with a data stream management system has been discussed with the event propagation, filtering, and aggregation component. To developed adaptation which makes it easy to interact with the monitoring system	Still the dynamicity is not reached, and enhancements are going on. New architecture are still in demand that can handle the dynamicity. Prediction mechanism has to be analysed to deal with such situation
5	Mian et al. [26]	The cost model has been discussed, which balances resource costs and penalties from SLAs if the SLA's are violated	Usage of static provisioning to provide an initial configuration and then moving to the concepts of dynamic enhancement is yet to be analysed
6	Ayad and Dippel [4]	Continuously check the availability of the virtual machines and automatically intervene in the case of VM failure If agent report that the machine is no longer healthy (corrupt or intrusion) to run, the monitor will destroy those machine, rolling back to the nearest healthy backup available and restart again. Destroying, recovering and restarting the VM should not take more time	Agents has to be made intelligence using other Machine Learning Techniques to get better results Needs to add more functionality in the agents and monitoring systems

(*continued*)

Table 1. (*continued*)

Sr.No	Authors	Objectives and methodologies	Conclusion and future directives
7	Li [23]	The Systematic Literature Review (SLR) method was employed to collect applicable suggestions to investigate the Cloud services evaluation turn by turn The time to time collection of evidences are used to make updates to the knowledge to focus upon new research areas	The metrics can be made based upon the following points The data from SLR will be stored into structured database in support of the services of cloud evaluation methodology to develop superior evaluation metrics
8	Da Cunha Rodrigues [12]	Survey paper which describes various monitoring techniques	Monitoring has to be achieved without compromising application performance and SLA's Integrating different cloud monitoring techniques together When specific requirement will come, it is either negatively or positively affected by other requirements, thus balancing among cloud monitor requirements is a challenging and important trend
9	Hill and Humphrey [17]	Create clusters of machines on-demand and use them for small to medium scale computational problems.	Root cause analysis: techniques able to derive the causes of the observed phenomena, spotting the right thread in the complex fabric of the Cloud infrastructure. Root cause analysis here indicates the primary factor which results into the failure of the system
10	Aceto [1]	Survey Paper which highlights that monitoring is required at both CSP and as well as the client side too	Cross Layer Monitoring: Consumers and Providers make their decisions based on a limited horizon. Both of them has to be considered With Cloud monitoring requirements also focus on minimizing the related resource/ energy consumption and monitoring of Federated Clouds is also a challenge
11	Botta et al. [7]	The workload modelling and generation has been discussed. Different contributions have been provided in terms of studies of real and synthetic workloads.	An important challenge is about the workload generators specifically designed for Cloud scenarios (with adaptability) and it should give the correct value if used for analysis of results

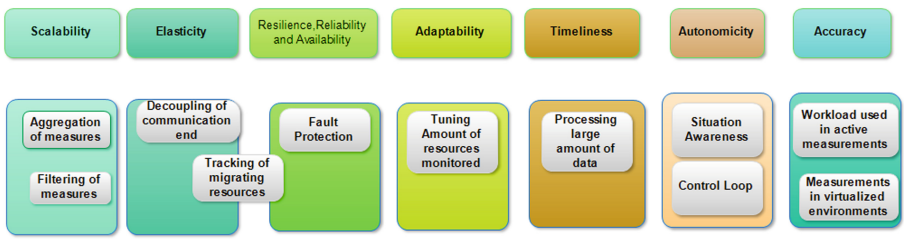


Fig. 1. Cloud monitoring necessity

Table 2. The Literature survey on prediction in to the cloud computing for efficient resource utilisation

Sr.no	Title authors	Objectives and methodology	Conclusion and future directions
1	Kousiouris [21]	It predicts the anticipated user behaviour (Behavioural level). Patterns are identified through a time series analysis	The potential usage of Support Vector Machines has to be analysed As it can perform better than ANN in various cases
2	Menascé and Almeida [25]	Analyse the customer behavioural pattern for website's workload characterization. Also make use of Customer Behaviour Model Graphs for calculating different metrics in order to find workload	To find the accuracy of these metrics
3	Almeida [2]	Different steps involved in capacity planning are discussed. This paper provides base for different activity of workload prediction	Different Tools like Matlab etc. can be used for forecasting, planning, analysis of work load depending on the applications type and its usages
4	Rimal et al. [28]	The author compare different cloud system on the basis of architecture, virtualization, storage, load balancing, interoperability, programming framework, security etc.	Which scheduling algorithms are suitable to which kind of environment? Is still an open research to be discussed among the researchers
5	Huang et al. [18]	To capture the relationship between the workload and the performance metrics. It is possible to ensure that performance of applications is above a minimum threshold. so the SLA violations can be avoided. To include the domain knowledge to model the application behaviour	Some controllers such as fuzzy controllers are based on the rule based approaches. The rules extraction is not easy for the resource management. The ability of the controllers depends on the defined rules and the rule based approaches do not have the learning capability. High availability is required
6	Buyya et al. [9]	The future demand of applications should be predicted accurately in a way that the resources manager is able to reallocate resources before the workload changes occurs	They cannot extract all useful patterns whose length is less/more than the fixed length. Choosing the length of the pattern (the length of the sliding window) for different regions of workloads is one of the most important challenges in these methods
7	Singh and Chana [29]	Different types of resources which include physical resources such as compute, memory, storage, servers, processors and networking are allocated to cloud applications were discussed	Most of the existing methods focus on one or two resources and ignore the correlation between resources. Researchers could investigate the correlation between these resources and provide more reasonable results for the resource manager
8	Urgaonkar [32]	Researchers could develop the new prediction approaches based on both of the reactive and the proactive methods. The proactive prediction methods should be able to extract all access patterns correctly	The reactive provisioning methods react to the surge of fluctuations or the deviation from the expected behaviour. They allocate the additional resources according to the workload increase to prevent SLA violation. Timeliness is the issue here
9	Buyya et al. [8]	Historical executions details (Statistical data) will be used for prediction of resources selection for the workload assigned	The market oriented principles for supply and demand of the resources should also be considered

(*continued*)

Table 2. (*continued*)

Sr.no	Title authors	Objectives and methodology	Conclusion and future directions
10	Ullrich and Lassig [31]	Make use of the pattern extracted from previous executions. Three different categories of load balancing: black box, grey box and white box were discussed	Predict the necessary resource adaption in real time if not even in advance. Resource Consumption can be applied based on the type of application
11	da Silva Dias [13]	Made use of monitoring agents for self-configuration	Self-Adaptive Capacity Management: Monitor and will respond to certain conditions that is overload or underutilisation of the resources at run time has to be analysed
12	Amiri and Mohammad-Khanli [3]	Survey Paper	The new approaches should be able to extract all the behavioural patterns of workloads independent of the fixed pattern length Capabilities of online learning has to be analyzed, able to identify the interesting trends or patterns of the workload variations. Researchers could develop the new prediction approaches based on both of the reactive and the proactive methods

prediction and resource monitoring [22]. The Literature survey on Monitoring in to the cloud computing for efficient resource utilization are as follow (Tables 1 and 2).

Fig. 1 shows [19] the relationship between the cloud properties to the below mentioned key points as, scalability depends upon the aggregation of measures and filtering of measure etc.

2 Prediction

There are various case studies [20,25] that indicates that the workload prediction plays very important role for any company. Basic Steps Required for Workload Prediction [2] and understanding the environment where the workload has to be executed, characterize the workload based upon its availability of resources or capacity planning, behaviour pattern etc., are key features for the prediction mechanism. The next step is to identify the parameter of the workload modelling, which depends upon the type of applications [2]. Lets now analyze the literature review of various prediction techniques.

3 Combining Monitoring and Prediction Techniques [10]

The above Fig. 2 indicates the connectivity between the monitoring and prediction mechanism to the cloud scenarios, their relationship has been identified [37] Prioritization Engine. As we can have more task aligned in the message queue in the cloud computing for different clients asking for the same resources in such scenario Business policies defined by the MSP (Managed Service Provider) helps

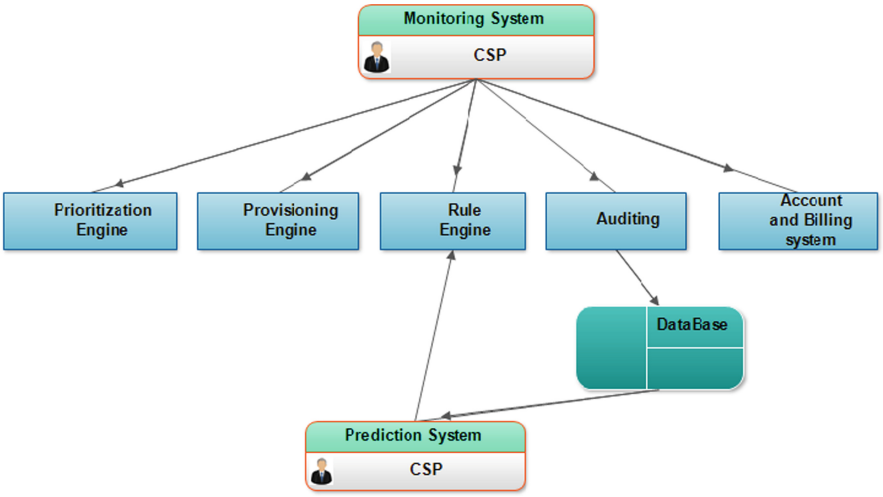


Fig. 2. The combined architecture

to identifying the requests whose execution should be prioritized with respects to the services that they want, in case of resource contentions [15]. The rules engine evaluates the data captured by the monitoring system [33]. Rules engine and the operational policy is the key to guaranteeing SLA agreements.

Monitoring System. Monitoring system collects the defined metrics in SLA. These metrics are used for monitoring resource failures, evaluating operational policies as well as for auditing and billing work. Monitoring system have to interact [6] with the other systems to optimise the its objective if careful usages of the resources available at IaaS of cloud.

Auditing. The adherence to the predefined SLA needs to be monitored and recorded. It is essential to monitor the values of SLA, as any noncompliance leads to strict penalties.

Prediction System: From auditing, the prediction model can be derived, which can be used to predict the resource consumption into the cloud and prediction can be merged with the Machine learning capabilities to increase its effectiveness.

Accounting/Billing System: Based on the payment model and metering The outcome of this model will predict the accurate resources usage for the specific type of work load and the problems that arises because of the under provisioning and over provisioning can be avoided.

4 Conclusion

In this survey paper we have highlighted the concepts of monitoring and prediction, which are an essentials for cloud computing environment, as the IaaS

gives us a vision of infinite pool resources and managing such a huge resources while serving at local level as well as at remote level site is a tedious task and can be handled efficiently if the mechanism of monitoring and prediction concepts should be mapped to the cloud environment. The algorithms related to these two can be merged with the techniques of mathematical modelling, artificial intelligence and machine learning for better accuracy of results and analysis. The model which has been discussed at the concluding portion in this paper will allow the researchers to impose the techniques to implement monitoring and prediction at the correct position with respect to its associated attributes of the cloud computing environment.

References

1. Aceto, G., Botta, A., De Donato, W., Pescapè, A.: Cloud monitoring: a survey. Comput. Netw. **57**(9), 2093–2115 (2013)
2. Almeida, V.A.F.: Capacity planning for web services techniques and methodology. In: Calzarossa, M.C., Tucci, S. (eds.) Performance 2002. LNCS, vol. 2459, pp. 142–157. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45798-4_7
3. Amiri, M., Mohammad-Khanli, L.: Survey on prediction models of applications for resources provisioning in cloud. J. Netw. Comput. Appl. (2017)
4. Ayad, A., Dippel, U.: Agent-based monitoring of virtual machines. In: 2010 International Symposium in Information Technology (ITSim), vol. 1, pp. 1–6. IEEE (2010)
5. Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Future Gener. Comput. Syst. **28**(5), 755–768 (2012)
6. Bose, S.K., Sundarrajan, S.: Optimizing migration of virtual machines across datacenters. In: International Conference on Parallel Processing Workshops, ICPPW 2009, pp. 306–313. IEEE (2009)
7. Botta, A., Dainotti, A., Pescapé, A.: A tool for the generation of realistic network workload for emerging networking scenarios. Comput. Netw. **56**(15), 3531–3547 (2012)
8. Buyya, R., Broberg, J., Goscinski, A.M.: Cloud Computing: Principles and Paradigms, vol. 87. Wiley, New York (2010)
9. Buyya, R., Calheiros, R.N., Li, X.: Autonomic cloud computing: open challenges and architectural elements. In: 2012 Third International Conference on Emerging Applications of Information Technology (EAIT), pp. 3–10. IEEE (2012)
10. Chen, H., Fu, X., Tang, Z., Zhu, X.: Resource monitoring and prediction in cloud computing environments. In: 2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), pp. 288–292. IEEE (2015)
11. Clayman, S., Galis, A., Mamatas, L.: Monitoring virtual networks with lattice. In: 2010 IEEE/IFIP Network Operations and Management Symposium Workshops (NOMS Wksps), pp. 239–246. IEEE (2010)
12. Da Cunha Rodrigues, G., Calheiros, R.N., Guimaraes, V.T., dos Santos, G.L., de Carvalho, M.B., Granville, L.Z., Tarouco, L.M.R., Buyya, R.: Monitoring of cloud computing environments: concepts, solutions, trends, and future directions. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing, pp. 378–383. ACM (2016)

13. da Silva Dias, A., Nakamura, L.H.V., Estrella, J.C., Santana, R.H.C., Santana, M.J.: Providing IaaS resources automatically through prediction and monitoring approaches. In: 2014 IEEE Symposium on Computers and Communication (ISCC), pp. 1–7. IEEE (2014)
14. Dillon, T., Wu, C., Chang, E.: Cloud computing: issues and challenges. In: 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), pp. 27–33. IEEE (2010)
15. Gor, K., Ra, D., Ali, S., Alves, L., Arurkar, N., Gupta, I., Chakrabarti, A., Sharma, A., Sengupta, S.: Scalable enterprise level workflow and infrastructure management in a grid computing environment. In: IEEE International Symposium on Cluster Computing and the Grid, CCGrid 2005, vol. 2, pp. 661–667. IEEE (2005)
16. Hasselmeyer, P., d'Heureuse, N.: Towards holistic multi-tenant monitoring for virtual data centers. In: 2010 IEEE/IFIP Network Operations and Management Symposium Workshops (NOMS Wksps), pp. 350–356. IEEE (2010)
17. Hill, Z., Humphrey, M.: A quantitative analysis of high performance computing with Amazon's EC2 infrastructure: the death of the local cluster? In: 2009 10th IEEE/ACM International Conference on Grid Computing, pp. 26–33. IEEE (2009)
18. Huang, D., He, B., Miao, C.: A survey of resource management in multi-tier web applications. IEEE Commun. Surv. Tutor. **16**(3), 1574–1590 (2014)
19. KaurSahi, S., Dhaka, V.S.: A review on workload prediction of cloud services. Int. J. Comput. Appl. **109**(9), 1–4 (2015)
20. Kohavi, R., Longbotham, R.: Online experiments: lessons learned. Computer **40**(9), 103–105 (2007)
21. Kousiouris, G., Menychtas, A., Kyriazis, D., Gogouvitis, S., Varvarigou, T.: Dynamic, behavioral-based estimation of resource provisioning based on high-level application terms in cloud platforms. Future Gener. Comput. Syst. **32**, 27–40 (2014)
22. Li, A., Yang, X., Kandula, S., Zhang, M.: CloudCmp: comparing public cloud providers. In: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, pp. 1–14. ACM (2010)
23. Li, Z., Zhang, H., O'Brien, L., Cai, R., Flint, S.: On evaluating commercial cloud services: a systematic review. J. Syst. Softw. **86**(9), 2371–2393 (2013)
24. Mell, P., Grance, T., et al.: The NIST definition of cloud computing (2011)
25. Menascé, D.A., Almeida, V.A.F.: Challenges in scaling e-business sites. In: Interantional CMG Conference, pp. 329–336 (2000)
26. Mian, R., Martin, P., Vazquez-Poletti, J.L.: Provisioning data analytic workloads in a cloud. Future Gener. Comput. Syst. **29**(6), 1452–1458 (2013)
27. Nida, P., Dhiman, H., Hussain, S.: A survey on identity and access management in cloud computing. Int. J. Eng. Res. Technol. **3**(4) (2014)
28. Rimal, B.P., Choi, E., Lumb, I.: A taxonomy and survey of cloud computing systems. In: INC, IMS and IDC, pp. 44–51 (2009)
29. Singh, S., Chana, I.: QoS-aware autonomic resource management in cloud computing: a systematic review. ACM Comput. Surv. (CSUR) **48**(3), 42 (2016)
30. Turab, N.M., Taleb, A.A., Masadeh, S.R.: Cloud computing challenges and solutions. Int. J. Comput. Netw. Commun. **5**(5), 209 (2013)
31. Ullrich, M., Lassig, J.: Current challenges and approaches for resource demand estimation in the cloud. In: 2013 International Conference on Cloud Computing and Big Data (CloudCom-Asia), pp. 387–394. IEEE (2013)
32. Urgaonkar, B., Shenoy, P., Chandra, A., Goyal, P., Wood, T.: Agile dynamic provisioning of multi-tier internet applications. ACM Trans. Auton. Adapt. Syst. (TAAS) **3**(1), 1 (2008)

33. Von Halle, B.: Business Rules Applied: Building Better Systems using the Business Rules Approach. Wiley Publishing, New York (2001)
34. Wang, C., Schwan, K., Talwar, V., Eisenhauer, G., Hu, L., Wolf, M.: A flexible architecture integrating monitoring and analytics for managing large-scale data centers. In: Proceedings of the 8th ACM International Conference on Autonomic Computing, pp. 141–150. ACM (2011)
35. Whiteaker, J., Schneider, F., Teixeira, R.: Explaining packet delays under virtualization. ACM SIGCOMM Comput. Commun. Rev. **41**(1), 38–44 (2011)
36. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. J. Internet Serv. Appl. **1**(1), 7–18 (2010)
37. Zhang, W., Song, Y., Ruan, L., Zhu, M.-F., Xiao, L.-M.: Resource management in internet-oriented data centers. Ruanjian Xuebao/J. Softw. **23**(2), 179–199 (2012)

Experimenting with Energy Efficient VM Migration in IaaS Cloud: Moving Towards Green Cloud

Riddhi Thakkar(✉), Rinni Trivedi, and Madhuri Bhavsar

Institute of Technology, Nirma University, Ahmedabad 382481, Gujarat, India
{15mcen27,14mcei28,madhuri.bhavsar}@nirmauni.ac.in

Abstract. Increasing demand for Cloud infrastructure and services leads to the challenges for management and maintenance of large data Center. Data center is fully equipped with huge number of resources. Those resources consumes energy in spite of their partial or full utilization. As a result data center consumes lots of energy, which in turn increases the total cost of operation and carbon footprint in environment. These concern leads to "Green Computing", i.e. to reduce total operational cost, Carbon Footprint in environment and efficient usage of the computing resources. In data center main processing element is virtual machine (VM), which is an instance of computing and storage resources, handles computational processes. Hence, it is important to reduce energy consumed by VM. As the workload distribution is varying in data Center as per the need, the number of VMs configured in the host are uneven, but host consumes maximum energy every time, irrespective of the workload. This leads to wastage of computational resources. This paper is intended to analyze such issues and specifically prove an algorithm which, significantly reduces energy consumption in data center, while ensuring SLA, when VM is in migration from one host to another in the data center.

Keywords: Energy efficiency · Green cloud · Vm migration

1 Introduction

Cloud computing is a computing model which provides access to shared pool of resources on demand, which can be swiftly provided and released with less management efforts and interaction of service provider and charged as much use [2]. It provides high availability and facilitating some important characteristics: Elasticity, on demand self service, Resource Pooling, and Network access. Cloud Computing Supports three service models IaaS, SaaS and PaaS, and uses virtualization technology to deliver scalability and elasticity to IT services. This facilitates to avoid over-provisioning and under-provisioning and to cater a new business utilities based on new ways of operating [3]. Cloud Computing is

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 56–65, 2018.
https://doi.org/10.1007/978-3-319-73712-6_6

widely used, but Resource Utilization, Security, Portability, and Interoperability are still major issues to deal with.

Computing resources inside data center consumes huge amount of energy, produces thousands of ton carbon dioxide and harms the atmosphere [5]. Energy is also required for functioning of the cooling system. Extra 0.5 W is required by the computing resources for each watt of power consumption for cooling [9]. According to [1] if any solution is not applied to reduce carbon emission by the data center, then it will be triple 2020 as compared to 2002. As indicated by the European Union, for keeping global temperature grow below 2°, it is necessary to drop the carbon emission up to 15% to 30% [11]. So, Consumption of Energy by Cloud Infrastructure as well as Carbon dioxide emission by that have been an important environment consideration. This will lead to use the Eco inviting computing called "Green Cloud" for minimizing the computing cost and to decrease the environmental destruction. Green means Environment-friendly and in other words utilization of resources.

Energy Efficiency can be achieved in Cloud Computing by providing some characteristics [9]: Improved Resource Utilization, Portability: VMs can be migrated to other physical host to reduce energy consumption, Elasticity, upgrading the running time of algorithms in application, DVFS and Virtualization of resources. In DVFS, the amount of voltage given to resource is decreased, when it is not completely utilized. The disadvantage is, it cannot apply to any resource of the cloud other than CPU, and frequency/voltage can be set to a limited number of states. Virtualization allows the user to create multiple VMs on single hardware and to run different OS on it. It provides heterogeneity, improves performance isolation and fault isolation by migrating VMs from one host to other host by Hot or Cold migration. In summary, it is necessary to utilize the cloud resources optimally, by migrating and consolidating VMs in less number hosts. This will decrease the carbon footprint and energy consumed by cloud resources.

2 Domain Analysis

Sherif et al. [4] investigated various techniques for migration and various stages in that techniques. VM migration parameters are classified into two parts: Static and Dynamic, based on their effect on VM migration. Static effect parameter has an unavoidable migration overhead. Dynamic effect parameter influences only the transfer of process, which are typically associated with the VM specification and its hosted application. The parameters that are strongly affecting the VM migration behavior are Page Dirty Rate and Link Speed. The author concluded that migration in XEN cannot escalate with a high-speed link.

Anton et al. [6] specified some key issues that must be undertaken for energy efficiency and performance. VM allocation policy is divided into two parts: The submission of a request for VM provisioning. For VM provisioning, changes are done in Best Fit Decreasing Algorithm. Optimization of VM allocation is performed in two steps: 1. To choose the VM for migration, 2. To choose an appropriate host for that VM. Author discussed RC, MMT, and highest potential

growth polices for choosing VM for migration. Paper shows, Minimum Migration policy perform less VM migrations and decrease SLA violations than other policies evaluated for simulation scenario.

Akshat et al. [13] have analyzed server workload of the large data center. To save energy, author experiment many attributes for the long and short term and conclude, there is an opportunity for saving power with help of off-peak metrics for categorizing application. While doing consolidation, if care is not taken then there will be possibilities of capacity violations. Two new consolidation approaches proposed: Peak Clustering based Placement (PCP) and Correlation Based Placement (CBP). These approaches use two metrics, first is an off-peak metric for categorizing and second is to guarantee that peaks don't cross limits. It is concluded, PCP provides better energy saving, decreased SLA violations and efficient load balance among active server.

Anton et al. [8] have investigated dynamic VM consolidation and single VM migration problem. In modern applications, because of variability in workload, VM placement is optimized online. For dynamic VM consolidation, author proposed some techniques, Median Absolute Deviation, Robust Local Regression, Interquartile Range, and Local Regression. These techniques use historical workload [7], generated from online services and web applications [4]. The goal of these techniques is to keep CPU utilization between upper and lower utilization threshold. For VM selection also, several policies discussed: MMT, RC, and Maximum Correlation (MC). From experiment, author shows that IRQ with MMT provides better energy reduction by reducing VM migration and SLA violations.

3 Energy Efficiency in IAAS Cloud

In data center, main processing element is VM, which is an instance of computing and storage resources, and does computational processes. So, VM plays important role in reducing energy consumption in data center. VMs are collected from an unutilized host. These VMs are consolidated into lesser number of host, and other hosts are switched off by migrating their load on active hosts. Hence, less number of host are active and energy is consumed by them only. In that way, overall energy consumption can be reduced in data center.

3.1 Proposed Architecture

Proposed architecture is intended for IAAS service. User submits a request for resource provisioning. Accordingly, SLA's are negotiated between Cloud Service Provider (CSP) and user. CSP does Service Scheduling and manages the account of the user. Monitoring kit is a heart of the system. It monitors power consumed by host and decides which host is over utilized or underutilized. It chooses VMs from over utilized host and consolidates them in less number of host by doing migration. VMs are configured on hosts. Each host is characterized by utilization of Million Instruction per Second (MIPS), Storage Capacity, network Bandwidth (BW) and RAM. User submits a request for provisioning

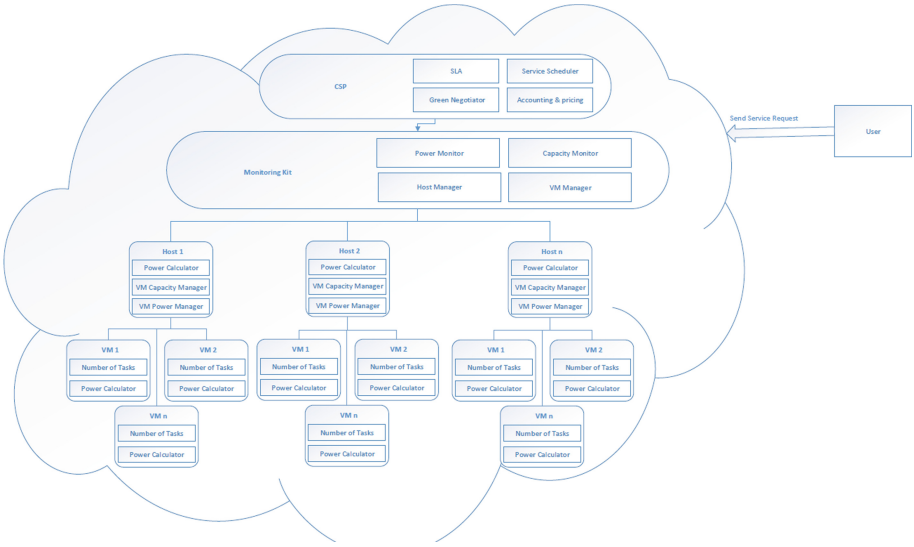


Fig. 1. Proposed architecture

of VMs, are allocated according to requested amount of MIPS, RAM, BW and Storage. While providing service to user CSP needs to take care of SLA.

3.2 Proposed System Model

Figure 2 shows, the snapshot of a system without implementation of energy efficient algorithm. In such system hosts are not fully utilized, there is less number of VMs are executing inside it then its capacity. When energy efficient algorithm is applied to the system, algorithm identifies underutilized host by comparing utilization parameter and collect the VMs from host. These VMs are migrated to the destination host which has enough resources for it and where VM consumes equal or less power. This process is repeated for all the VMs and are consolidated into less number of hosts. Hence, by migration of VMs, energy consumption of data center can be reduced.

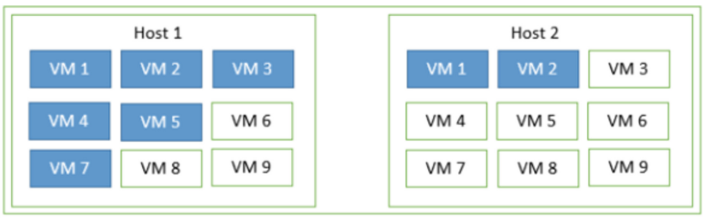


Fig. 2. Snapshot of system without implementation of energy efficient algorithm

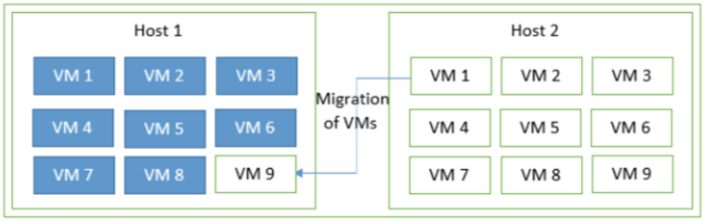


Fig. 3. Snapshot of system after execution of energy efficient algorithm

3.3 Descriptive Statistic: Interquartile Range

Interquartile Range is used to identify over utilized hosts. IQR is choosen, because it is less sensitive to highly skewed data. This method uses historical data of CPU utilization and provides maximum CPU threshold value. It finds the median of data. IQR is difference between median of data series after and before the median value.

$$IQR = m3 - m1$$

$$MaxCPUThreshold = 1 - s * IQR$$

Here, s [8] is a safety parameter, which shows that how aggressive consolidation of VMs are done inside host.

4 Proposed Algorithm

To reduce energy consumption, VMs are needed to be collected from unutilized hosts. For this purpose, first over utilized hosts needs to be identified, IQR is used for it. Host with CPU utilization more than MaxCPUUtilization are over utilized host. For collecting VMs from host three methods are preferred: MMT, RC and Maximum Correlation (MC). In minimum migration time, VMs are selected which requires less time for migration. Migration time is calculated as amount of RAM used by VM divided by spare BW available for selected Host. In Random choice method, VM is selected randomly. In Maximum correlation, VMs are chosen based on maximum correlation of utilization of CPU with other VMs utilization of CPU. Multiple correlation coefficient is used for identifying a correlation between utilization of a CPU of VMs [8]. We use minimum migration time technique for collection of VM. After choosing VM, the host is switched off, if it is empty. Under utilized hosts are collected by excluding over utilized and switched off hosts. All the VMs from such host are collected and consolidate into less number of hosts.

VMs are required to place in appropriate hosts as per polices and availability of resources inside host. VMs are sorted in decreasing order of their utilization of CPU. Power consumption of all the VMs are compared and maximum value is assigned to minPowerConsumption. For every VM, hosts are checked for required

Algorithm 1. VM Optimization

```
H: Host List, h_i: Each host inside Host List
for each (h_i ∈ H) do
if (h_i.CPUUtilization > MaxCPUThreshold) then
    Add VmsToMigrateFromOverloadedHost (h_i) to OverLoadedVmList
end if
Add OverLoadedVmList to VMMappingList
excludeHosts.add(OverUtilizedHosts)
excludeHosts.add(SwitchOffHosts)
excludeHosts.add(HostsInMigration)
end for
for each (h_i ∈ H) do
if (excludeHosts.contain(h_i)) then
    continue()
end if
if (utilization > 0 AND utilization < minUtilization) then
    minUtilization = utilization;
    underUtilizedHost = host;
    Add vm.underUtilizedHost to VMMappingList
    host.deallocateVM()
end if
end for
return VMMappingList
```

Algorithm 2. VM Placement

```
H: List of Switched Off and unutilized Host
vmList.sortDecreasingUtilization()
minPowerConsumption = MAX()
for each (v_i ∈ VMMappingList) do
allocatedHost = NULL
for each (h_i ∈ H) do
if (h_j.hasResources(v_i)) then
    powerReqByVm = estimatePower(h_j, v_i)
    if (powerReqByVm < minPowerConsumption) then
        allocatedHost = h_j
        minPowerConsumption = powerReqByVm
    end if
    if (allocatedHost <> NULL) then
        allocationOfHost.add(vm, allocatedHost)
    end if
end if
end for
end for
return allocation
```

resources. If the host contains sufficient resources for VM, then the power consumption of VM is compared with minPowerConsumption, and if it is less, then VM is allocated to that host. This way by analyzing the usage statistics, consolidation of VMs in less number of hosts is achieved. In result, total energy consumed by the data center is reduced, which is shown further in results.

5 Performance Analysis

The proposed technique is implemented on Private Cloud Infrastructure. For evaluation of the results generated after deploying an algorithm on the cloud, sufficient amount of resources are considered. It was very challenging to, because of it's complexity. For performing experiment dual-core HP ProLiant ML110 G4 and HP ProLiant ML110 G5 are used, with MIPS ranging from 1860 to 2660, RAM 4096, storage 10000000 and BW 10000000. VMs characterized with MIPS range from 500 to 2500, RAM range from 613 to 1740, Storage 2500 and BW 100000.

Power Model. By recent study [10,12], in data center CPU has linear power to frequency relationship with DVFS application. It also shows that the server running in idle mode approximately consumes 70% of the energy used by CPU working at full speed. So, to avoid this, the load is consolidated in less number of hosts. The model of power be defined by [6],

$$p(u) = k * P_{max} + (1 - k) * P_{max} * u$$

Here, Pmax: Max power consumption when server is utilized fully, k: part of power consumed when the server is idle mode, u: utilization of CPU. As workload is not stable, Utilization of CPU change over the time. So, it is a function of time, defined as u(t). Total Energy consumption E, defined as [6],

$$E = \int_{t0}^{t1} P(u(t))dt$$

Power consumption according to utilization is shown in the Table 1 [8],

Table 1. Power consumption according to CPU utilization [8]

Server	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
HP ProLiant G4	86	89.4	92.6	96	99.5	102	106	108	112	114	117
HP ProLiant G5	93.7	97	101	105	110	116	121	125	129	133	135

5.1 Evaluation of Result

For implementation, 3 physical hosts are taken. Results shows that, the hosts consumes an equal amount of energy during whole time frame if it is ideal or fully utilized. This is results in wastage of resources and increase in energy consumption. By implementing proposed algorithm, resource utilization can be increased and energy consumption can be reduced.

The results after implementation of algorithm shows that, Energy consumption of hosts is decreased by migrating VMs from unutilized hosts and consolidating them in less number of hosts. The host does not consume energy when it is not utilized.

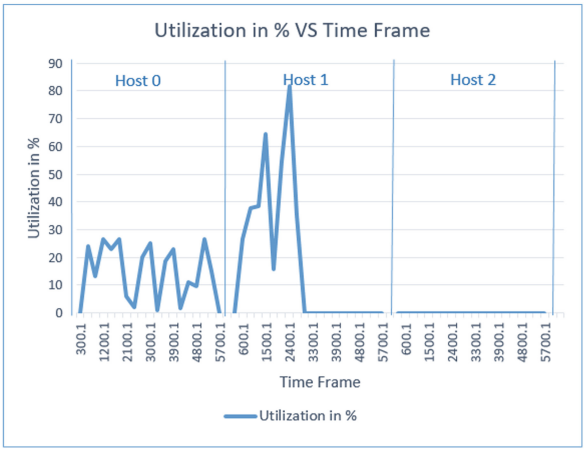


Fig. 4. Time frame VS Utilization without implementation of energy efficient algorithm

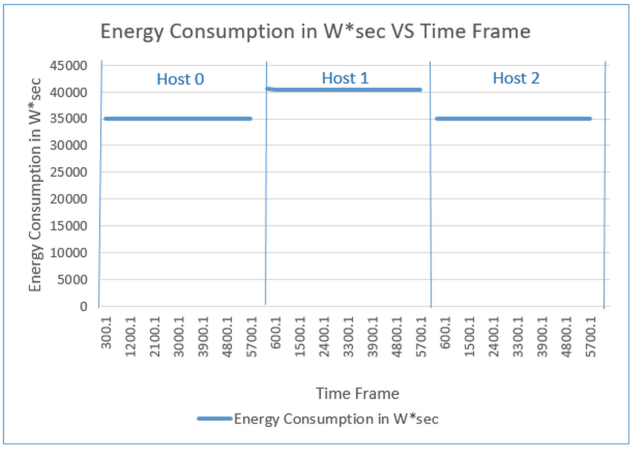


Fig. 5. Time frame VS Energy consumption without implementation of energy efficient algorithm

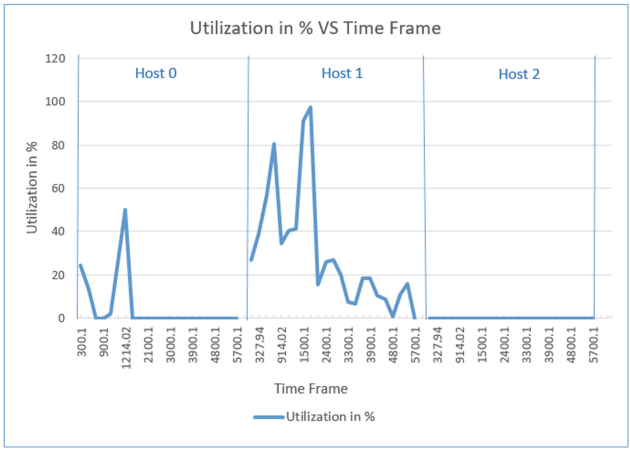


Fig. 6. Time frame VS Utilization after implementation of energy efficient algorithm

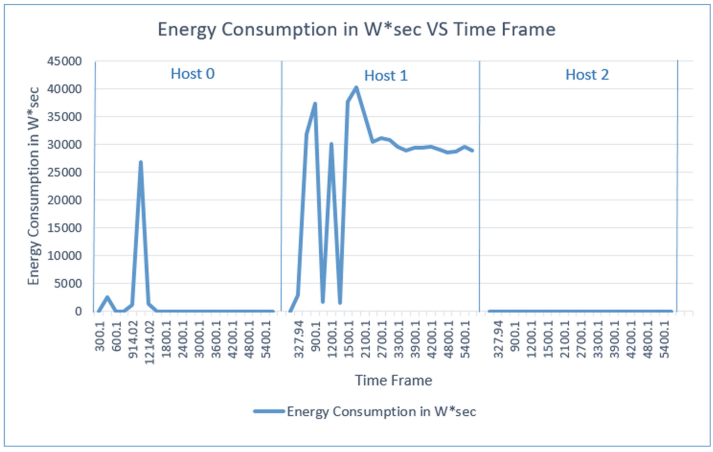


Fig. 7. Time frame VS Energy consumption after implementation of energy efficient algorithm

6 Conclusion and Future Work

Energy Efficient Computing reduces Carbon Footprint in the environment and increases return of investment in cloud resources. Proposed technique performs energy efficient computing at VM level and increases utilization of resources, which in turn reduces active number hosts and energy consumed by the host. This way energy consumed by the data center is decreased.

The research work is intended to analyse energy consumption of host machine in cloud and perform appropriate VM migration in case of overconsumption of

energy. As hardware resources are majorly handled in IaaS, PaaS and SaaS services for energy monitoring are kept as future work.

References

1. Cloud computing and sustainability: The environmental benefits of moving to the cloud (2010)
2. Building return on investment from cloud computing - executive summary (2017)
3. NIST cloud computing program (2017)
4. Akoush, S., Sohan, R., Rice, A., Moore, A.W., Hopper, A.: Predicting the performance of virtual machine migration. In: 2010 IEEE International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), pp. 37–46. IEEE (2010)
5. Beloglazov,A.: Energy-efficient management of virtual machines in data centers for cloud computing. Ph.D. thesis (2013)
6. Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Future Gener. Comput. Syst. **28**(5), 755–768 (2012)
7. Beloglazov, A., Buyya, R.: Energy efficient allocation of virtual machines in cloud data centers. In: 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid), pp. 577–578. IEEE (2010)
8. Beloglazov, A., Buyya, R.: Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurr. Comput. Pract. Exp. **24**(13), 1397–1420 (2012)
9. Beloglazov, A., Buyya, R., Lee, Y.C., Zomaya, A., et al.: A taxonomy and survey of energy-efficient data centers and cloud computing systems. Adv. Comput. **82**(2), 47–111 (2011)
10. Fan, X., Weber, W.-D., Barroso, L.A.: Power provisioning for a warehouse-sized computer. In: ACM SIGARCH Computer Architecture News, vol. 35, pp. 13–23. ACM (2007)
11. Garg, S.K., Buyya, R.: Green cloud computing and environmental sustainability. In: Murugesan, S., Gangadharan, G. (eds.) Harnessing Green IT: Principles and Practices, pp. 315–340. Wiley, London (2012)
12. Kusic, D., Kephart, J.O., Hanson, J.E., Kandasamy, N., Jiang, G.: Power and performance management of virtualized computing environments via lookahead control. In: 2008 International Conference on Autonomic Computing, ICAC 2008, pp. 3–12. IEEE (2008)
13. Verma, A., Dasgupta, G., Nayak, T.K., De, P., Kothari, R.: Server workload analysis for power minimization using consolidation. In: Proceedings of the 2009 Conference on USENIX Annual Technical Conference, p. 28. USENIX Association (2009)

Capacity Planning Through Monitoring of Context Aware Tasks at IaaS Level of Cloud Computing

Vivek Kumar Prasad(✉), Harshil Mehta, Parimal Gajre, Vidhi Sutaria, and Madhuri Bhavsar

Nirma University, Ahmedabad 382481, Gujarat, India
{vivek.prasad,15mcen12,15mcec12,15mcei28,madhuri.bhavsar}@nirmauni.ac.in
http://www.nirmauni.ac.in/

Abstract. Cloud Computing is the exercise of using a network of remote servers held on the Internet to store, manage, and process data which have the characteristics as an elasticity, scalability or scalable resource sharing managed by the resource management. Even the growing demand of cloud computing has radically increased the energy consumption of the data centres, which is a critical scenario in the era of cloud computing, hence the resources has to be used efficiently, which ultimately will minimise the energy. Resource management itself will get the data from resource monitoring and resource prediction for the smooth conduction of the tasks and its allocated resources. In this paper the monitoring mechanism in the cloud has been discussed and its results are used to trigger the prediction rule engine which provides the cloud service provider (CSP) to start allocating the resources in the efficient manner, even the concept of failure handling has been mentioned based upon the certain parameter which will also inform the CSP to handle the failure task and try to mitigate this and again re schedule the failed task.

Keywords: Cloud computing · Monitoring · Resource management
Resource prediction · Scheduling · Error handling

1 Introduction

Cloud computing [9] is a major force, which is changing the Information technology landscape and moving it entire data to the cloud for putting it into the remote location to store, manage and process the data using Internet.

Cloud Service Model: The services are classified based upon their functionality, i.e. Software as a service (SaaS) which is used to deliver the web applications, Platform as a Service (PaaS), to create or deploy application and services for user, Application test, development, integration and deployment type of services. Infrastructure as a service, provides services as rent storage, processing and communication using the concepts of virtual machine.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 66–74, 2018.
https://doi.org/10.1007/978-3-319-73712-6_7

In this paper we have proposed the monitoring of the cloud computing environment, in such a way so that the resources utilisation state will be observed continuously and the threshold value of the resources has to be identified and if the value reaches at the threshold, then for the efficient usage of the resources prediction (and its profiling) mechanism will be invoked, which will allow the CSP to make efficient usage of resources, so that the cloud can handle maximum number of requests, thus ultimately will lead to increase in the revenue. The context aware tasks profiling will make us sure that exactly how much resources the particular task will consume and its behaviour for the future request also. Any deviation from its normal behaviour will make the CSP aware of something wrong has happened in the cloud, which in turn will trigger the failure handling and its mitigation approach towards the handling of the failed tasks and again reschedule them.

2 Monitoring the Cloud Environment

2.1 Monitoring as an Essential Tool in Cloud Computing Environment

Monitoring [1] is important for both provider and consumers for managing and controlling the hardware and software infrastructure, it also provides the key performance indicators and information for both platforms and applications. It is also useful for capacity planning [12] where the estimation of the correct resources will improve the efficiency of the resource utility, which will help to meet the criteria of SLA management [14].

Proposed Algorithm. The above Table 1 indicates the execution time taken by various categories of scheduling algorithms based upon the varied availability (the number of Virtual Machines availability) of the resources at the cloud computing environment, Which indicates the scheduling algorithms performs better in different availability of resources, as cloud is dynamic in nature [18]. The resources are in terms of Virtual Machine.

Table 1. Execution time of 25 task

Algorithm	VM 10	VM 25	VM 50	VM 80
FCFS	2541.95	1782.58	677.45	401.01
MCT	430.20	314.6	252.8	236.8
MINMIN	339.2	270.8	240.3	244
MAXMIN	274.3	259.40	248.4	238.2
RR	1449.4	1158.5	930.95	924.57
DATA AWARE	2435.04	1221.65	1009.09	518.25

Algorithm 1. Dynamic Scheduling Mechanisum

```
Initialization
Let the current CPU, RAM is assumed as resources are available on cloud.
{
Initialize threshold value =50%
When new task arrives check threshold value(resources).
if Threshold value >current resource utilization then
   {
   below threshold value;
   schedule the task;
   }
else
   (Threshold values[i]<current resource utilization)
   {
   above threshold value;
   Redirect to check-pointing mechanism();
   }
end if
}
```

The Table 2 above indicates the resources (CPU and RAM) usage values at different interval of time while executing tasks in cloud computing environment. Figure 1 shows the result of the Table 2 where x axis indicates the time intervals and y axis indicates the utility of the resources.

While considering the CPU usage of Table 2 the observations indicates that one observation has value above 50%, and rest are below 50%, on the flipped side in case if number of readings are more than 50% utilization of CPU indicates that the rest of the available resources should be utilised wisely. Likewise same observations has to be noted down for the memory utilization also. Now resource

Table 2. Data observed during experimental

Date and Time	CPU Usage	RAM Usage
April 13 2017 10:14 AM IST	16.02	54
April 13 2017 10:20 AM IST	40.87	67
April 13 2017 10:25 AM IST	34.38	74
April 13 2017 10:30 AM IST	28.2	69
April 13 2017 10:35 AM IST	13.88	68
April 13 2017 10:40 AM IST	78.99	74
April 13 2017 10:45 AM IST	72.3	71
April 13 2017 10:50 AM IST	83.26	73
April 13 2017 10:55 AM IST	65.3	52
April 13 2017 11:00 AM IST	68.80	44

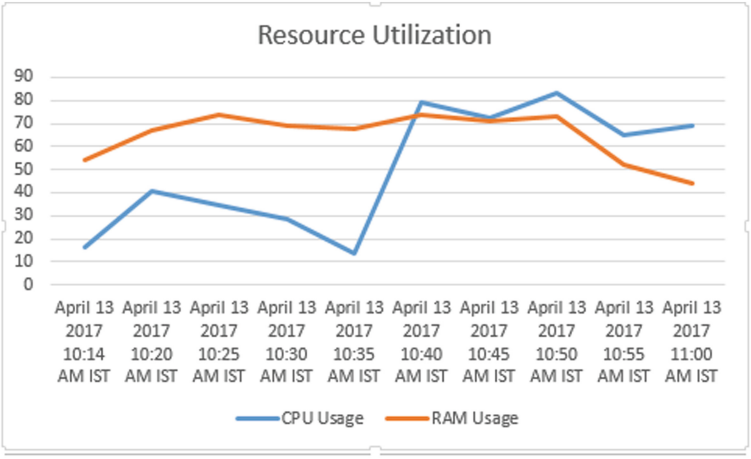


Fig. 1. Resources utilization

prediction for context aware workload module will be invoked which is explained in Sect. 2, So that the SLA will be maintained.

3 Resource Prediction for Context Aware Workload

3.1 Profiling in Cloud Computing for Context Aware Workload

Profiling is a mechanism through which the behaviours of the task execution can be recorded and can be used for the audit purpose. Profiling can be done in two ways, active profiling and passive profiling, where the active profiling is fine grain and the passive profiling is a coarse grain [13].

Here in this research paper for profiling we are considering only the context aware tasks, so that their total execution time and resource usage are determined [16], then these execution time has been divided into certain check points [4] and in every check points whatever resources has been consumed by the tasks are noted and a metric has been prepared, so that if the same task comes next time in future, the same metrics can be used to evaluate the performance of the task [16].

3.2 Proposed Algorithm

Figure 2 indicates the clustering of tasks based on their resource utilisation patterns and the experimentation has been done using weka. If any discrepancy occur due to the pattern mismatch (i.e. the stored pattern and the current pattern), then the error handling mechanism will be invoked, which is mentioned in Sect. 4 below.

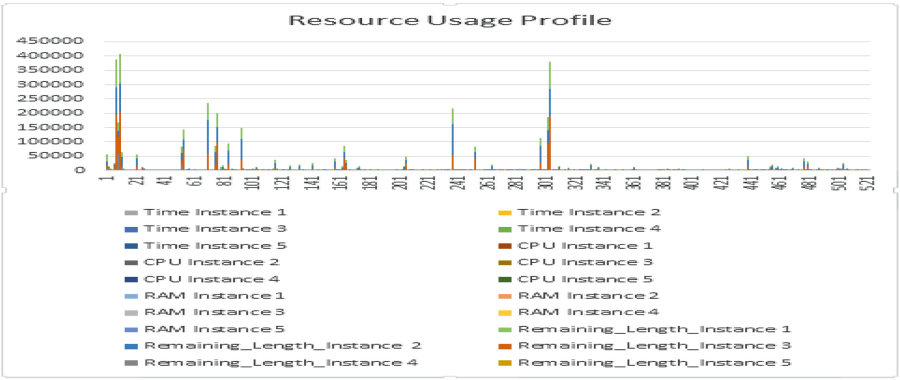


Fig. 2. Profiling of resources for checkpoints

Algorithm 2. Check pointing Mechanism

```
Divided whole task into 5 frame, each of 0.20.
compare with the previous data set by context aware mechanism
if Task match with previous data set task then
    {
    Scheduled the task to server;
    }
else {Task not match with previous data set task}
    {
    check for error handling;
    Redirect to error handling mechanism();
    }
end if
```

4 Error Handling and Mitigation

In this section we have highlighted the mechanism of the error handling and its mitigation techniques [15]. Error handling is the procedure of finding errors in the system. Error should be handle in dynamic way in cloud computing [8,11]. Error handling will also provide robustness and system availability against hardware and software errors in cloud [2]. The proactive method deals with recovery of fault in advance, whereas reactive method deals with recovery after the occurrence of error [3,5,6,10,17] Reactive mitigation techniques: Check-pointing, Restart, Replication, Job migration, Sguard, Retry, Task resubmission and Recue workflow [7].

4.1 Proposed Algorithm

In the given algorithm it classifies hardware and software errors and invoke appropriate mitigation technique for reduce the adverse effect of the error.

Algorithm 3. Error Handling Mechanism

```
Let assumed it will be hardware error or software error.
{
predict the error with error prediction;
if error code <500 then
    {
    software error;
    Mitigate software error by mitigation technique;
    }
else
    (error code>500)
    {
    Hardware error;
    Mitigate software error by mitigation technique;
    }
end if
Dynamic schedule task to the server;
go to next task;
}
```

Figure 3 mention about the execution time verses density (error), X- axis shown the execution time and Y- axis shown the Density, by this we can conclude that which type of error will occur (it has been mentioned in top right corner of figure as error names).

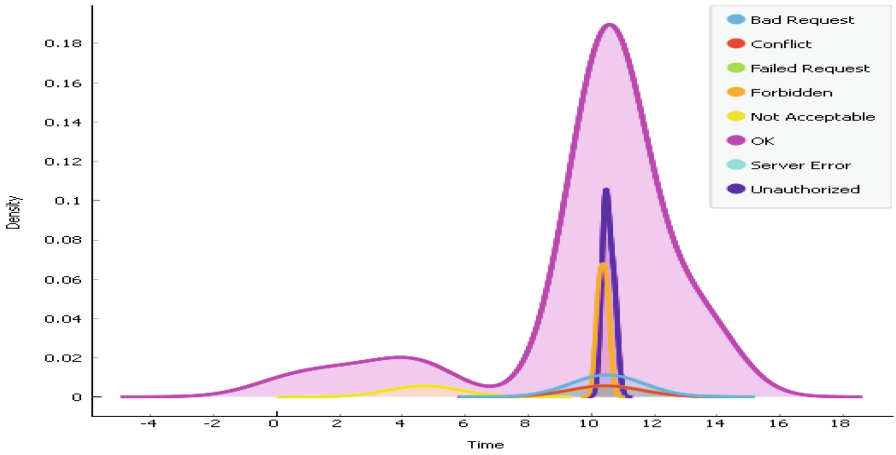


Fig. 3. Execution time Vs Density

5 Proposed Model

Proposed model is as shown in Fig. 4. Key items of the proposed model are as follow:

5.1 Guaranteed SLA Management

The agreement between the client and Cloud Service Provider (CSP) based on certain QoS parameter will be established and monitioring will be done based upon agreed SLA.

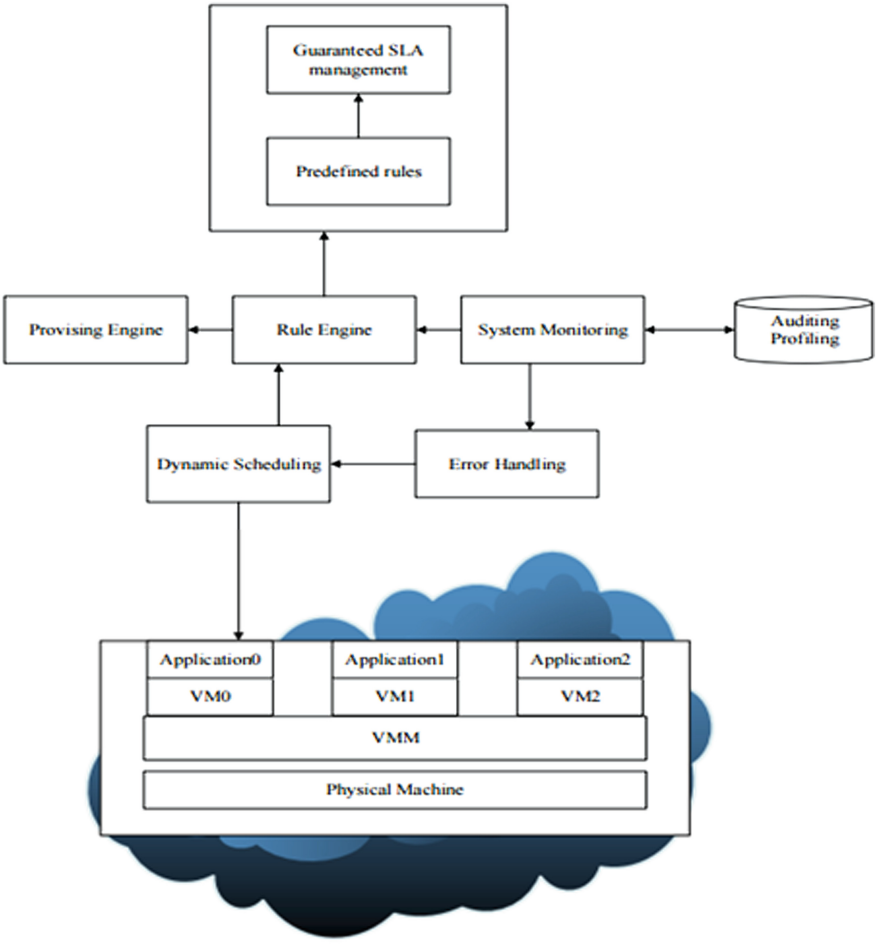


Fig. 4. Proposed model

5.2 Provisioning Engine

Functionality of provisioning engine is to enact according to set of steps known as provisioning plan. It is responsible for various requests from users and applications like to start an application, stop an application, halt an application, request for more resources and so on.

5.3 Rules Engine

Rules Engines functionality is to evaluate data captured by monitoring system based on operational policy. Operational policy defines different action sequence that should be triggered in incase of occurrence of an event. So Rules Engine and operational policy together provide the key to guaranteeing SLA under a self-healing system.

5.4 System Monitoring

Functionality of System Monitoring module is to collect information of different metrics which affect the performance of system and are defined in SLAs.

5.5 Auditing/Profiling

The attachment to the predefined SLA needs to be recorded and monitored. It is indispensable to monitor the compliance with SLA.

5.6 Dynamic Scheduling

As cloud is elastic in nature, the demand of the resources will always be fluctuating, so we require mechanism that will adapt to changing resources scenario.

5.7 Error Handling

As cloud is a on-demand network access to resources and software application, so there is chance that we can encounter the scenario where hardware or software may fail because of uncertainty, so we require some mechanism through which this errors can be handled and mitigated automatically without intervention of human being and after mitigation it should be reschedule back to the dynamic scheduler.

6 Conclusion and Future Work

In this research paper we have proposed an efficient capacity planning at IaaS level of cloud computing for context aware workload. To achieve this cloud monitoring concepts has been used and it has been incorporated to Hidden Markov model (HMM) to categorise the usage of resources at cloud. In critical state of

cloud resource usage profiling/auditing mechanism has been triggered. To handle the faulty scenarios where the resources are in peak demand has also been covered.

In this study only CPU and RAM are considered. However still there are other resources which are to be considered such as network, IO and so on. In our future works we will take these factors into consideration. We also will develop and built better energy efficient resource provisioning.

References

1. Aceto, G., Botta, A., De Donato, W., Pescapè, A.: Cloud monitoring: a survey. Comput. Netw. **57**(9), 2093–2115 (2013)
2. Agarwal, H., Sharma, A.: A comprehensive survey of fault tolerance techniques in cloud computing, pp. 408–413 (2015)
3. Bala, A., Chana, I.: Fault tolerance-challenges, techniques and implementation in cloud computing. IJCSI Int. J. Comput. Sci. Issues **9**(1), 1694–0814 (2012)
4. Bouteiller, A., Lemarinier, P., Krawezik, K., Capello, F.: Coordinated checkpoint versus message log for fault tolerant mpi. In: Proceedings of the 2003 IEEE International Conference on Cluster Computing, pp. 242–250. IEEE (2003)
5. Cheraghlou, M.N., Khadem-Zadeh, A., Haghparast, M.: A survey of fault tolerance architecture in cloud computing. J. Netw. Comput. Appl. **61**, 81–92 (2016)
6. Ganesh, A., Sandhya, M., Shankar, S.: A study on fault tolerance methods in cloud computing, pp. 844–849 (2014)
7. Jhawar, R., Piuri, V., Santambrogio, M.: Fault tolerance management in cloud computing: a system-level perspective. IEEE Syst. J. **7**(2), 288–297 (2013)
8. Kaur, P.D., Priya, K.: Fault tolerance techniques and architectures in cloud computing-a comparative analysis, pp. 1090–1095 (2015)
9. Mell, P., Grance, T., et al.: The nist definition of cloud computing (2011)
10. Mittal, D., Agarwal, N.: A review paper on fault tolerance in cloud computing, pp. 31–34 (2015)
11. Patra, P.K., Singh, H., Singh, G.: Fault tolerance techniques and comparative implementation in cloud computing. Int. J. Comput. Appl. **64**(14), 37–41 (2013)
12. Psoroulas, I., Anagnostopoulos, I., Loumos, V., Kayafas, E.: A study of the parameters concerning load balancing algorithms. IJCSNS Int. J. Comput. Sci. Netw. Secur. **7**(4), 202–214 (2007)
13. Ren, G., Tune, E., Moseley, T., Shi, Y., Rus, S., Hundt, R.: A continuous profiling infrastructure for data centers, Google-wide profiling (2010)
14. Shin, S., Kim, Y., Lee, S.: Deadline-guaranteed scheduling algorithm with improved resource utilization for cloud computing. In: 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), pp. 814–819. IEEE (2015)
15. Singla, N., Bawa, S.: Priority scheduling algorithm with fault tolerance in cloud computing. Int. J. 3(12) (2013)
16. Sotomayor, B., Keahey, K., Foster, I.: Combining batch execution and leasing using virtual machines. In: Proceedings of the 17th International Symposium on High Performance Distributed Computing, pp. 87–96. ACM (2008)
17. Tchana, A., Broto, L., Hagimont, D.: Approaches to cloud computing fault tolerance, pp. 1–6 (2012)
18. Zhong, H., Tao, K., Zhang, X.: An approach to optimized resource scheduling algorithm for open-source cloud systems. In: 2010 Fifth Annual ChinaGrid Conference, pp. 124–129. IEEE (2010)

ApEn-Based Epileptic EEG Classification Using Support Vector Machine

Hardika B. Gabani(✉) and Chirag N. Paunwala

Dr. R.K. Desai Marg, Opp. Mission Hospital, Athwalines, Surat, Gujarat, India
hardikagabani62@gmail.com, chirag.paunwala@scet.ac.in

Abstract. The ElectroEncephaloGram (EEG) signal plays an important role to identifying the disorder of epilepsy. Epilepsy is a neurological disorder which is an unexpected electrical disturbance of the brain. Due to which nerve cell activity in the brain becomes disrupted, causes people to have a "Seizure". Now a day researchers are working and focusing on an automatic analysis of EEG signal to classify the Epilepsy. The EEG signal recording system generate very lengthy data. So, classification of epilepsy seizure requires a time-consuming process. This paper proposes SVM (Support Vector Machine) based automatic epilepsy seizure classification system that uses ApEn (Approximation Entropy). ApEn is reducing the patient data size without any loss of patient data so; we can easily classify the epilepsy seizure. ApEn is statistical parameters that measure the current amplitude value of an EEG signal based on its previous amplitude value. In this paper, we measure sensitivity, specificity, and accuracy using SVM classifier. The overall values as high as 100% can be achieved using the proposed system to differentiate epileptic state (Seizure class) out of normal state (Non-seizure Class) using time domain method.

Keywords: ElectroEncephaloGram (EEG) signal
Epilepsy seizure classification · Approximate Entropy (ApEn)
SVM (Support Vector Machine)

1 Introduction

Electrical activity is occurring in different brain regions which are determined by the EEG signal and we can also determine the relative positions and strengths of it. The abnormal electrical activities fetched by using EEG are called epilepsy seizures. Approximately 50 million people have epilepsy seizure worldwide [1]. Possible causes of epilepsy include brain injury, metabolic disturbances, alcohol or drug abuse, brain tumors, and genetic disorders.

In the small time period, epileptic seizure can't be predicted in most of the cases. For classification purpose, continuous recording of the EEG is required. Sometimes EEG recording takes very large time duration. It may be up to one week or two weeks. As the traditional methods are monotonous and slow. In past few years, automated epilepsy seizure classification systems have been developed [2]. The proposed work is an automatic epileptic EEG classification system using SVM and feature extraction and reduction by using Approximate Entropy (ApEn).

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 75–85, 2018.
https://doi.org/10.1007/978-3-319-73712-6_8

As shown in below figure we give the EEG signal at the input side. ApEn technique [3] is used to extract the features of the signal. Extracted features are then apply to the classifier to classify seizures or non-seizures data (Fig. 1).

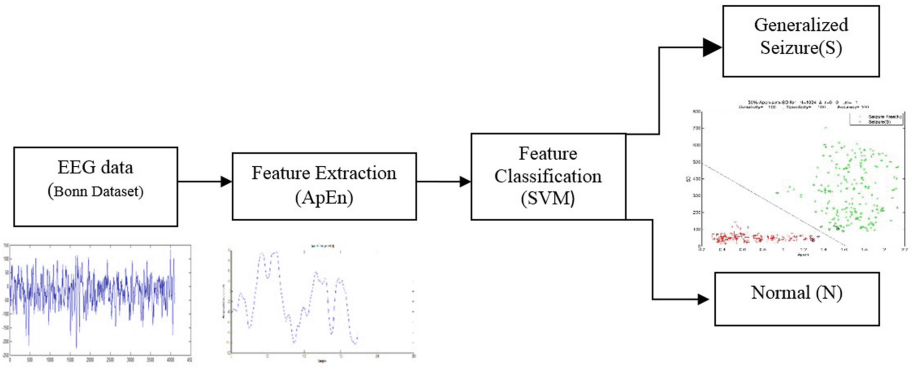


Fig. 1. Block diagram

Programmed examination and finding of epilepsy in view of EEG recordings is begun in the mid-1970s. Today, PC-based examination addresses two problems: Epilepsy seizure classification and EEG analysis. Many feature extraction techniques have been used for the classification of Epilepsy seizure. SVM (Support Vector Machine) based classification system for epilepsy seizure have been proposed by many researcher. The research based on nonlinear parameters has been found clinically fruitful for classification of Epilepsy seizure.

The Lyapunov exponent [4–6] provides significant details about changes in EEG activity in turn facilitating early detection of epilepsy. The correlation dimension [7] is useful to measure correlation which quantifies complex neural activity of human brain. During epileptic seizure, the value of ApEn has been found to exhibit strong relationship with synchronous discharge of large groups of neurons. The features obtained from complexity analysis and spectral analysis of EEG signals has been effectively used for diagnosis of epilepsy [8]. Recently, the ApEn (Approximate Entropy) [3] based methods have been developed for analyzing linear signals for classification of epileptic seizures in epilepsy seizure [9, 13]. The MEAN frequency parameter of IMFs has been proposed to discriminate well between seizure and seizure-free EEG signals. For classification between healthy and epileptic EEG signals, weighted frequency has been found to be some parameter [10]. Analysis of normal and epileptic seizure EEG signals by using area measured from the trace of analytical signal representation of Intrinsic Mode Function (IMF) has been proposed in [11]. The area parameter and mean frequency of IMFs computed using Fourier–Bessel expansion used for epileptic seizure classification in EEG signals [12]. Also, IMFs of EEG signals have been used for recognition of epileptic seizure [13].

2 Proposed Algorithm

2.1 ApEn (Approximate Entropy) Based Feature Extraction

An ApEn is a technique used to quantify the amount of regularity and the unpredictability of fluctuations over time-series data [3].

(1) Let EEG signal with N data points $X = [x(1), x(2), x(3), \ldots, x(N)]$.
(2) Let $x(i)$ be a subsequence of X such that $x(i) = [x(i), x(i+1), x(i+2), \ldots, x(i+m-1)]$ for $1 \leq i \leq N-m$, where m represents the number of samples used for the prediction.
(3) To reduce the noise, filter with level r is represented as, $r = k * SD$ for $k = 0, 0.1, 0.2, 0.3, \ldots, 0.9$
Where SD is the standard deviation of X.
(4) Let $\{\boldsymbol{x}(j)\}$ represent a set of subsequence's obtained from $\boldsymbol{x}(j)$ by varying j from 1 to N. Each sequence $\boldsymbol{x}(j)$ in the set of $\{\boldsymbol{x}(j)\}$ is compared with $\boldsymbol{x}(i)$ and, in this process, two parameters, namely, $C_i^m(r)$ and $C_i^{m+1}(r)$ are defined as follows:

$$C_i^m(r) = \frac{\sum_{j=1}^{N-m} k_j}{N-m} \tag{1}$$

Where,

$$k = \begin{cases} 1, & \text{if } |\boldsymbol{x}(i) - \boldsymbol{x}(j)| \text{ for } 1 \leq j \leq N-m \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$C_i^{m+1}(r) = \frac{\sum_{j=1}^{N-m} k_j}{N-m}$$

(5) Finally, we get Approximation Entropy,

$$\text{ApEn}(m, r, N) = \frac{\sum_{i=1}^{N-m} \ln(c_i^m(r))}{N-m} - \frac{\sum_{i=1}^{N-m} \ln(c_i^{m+1}(r))}{N-m} \tag{3}$$

Approximation entropy value extracted from the different size of data frames is shown in Table 1. Now as from the algorithm of ApEn, m is sample value varies from 1 to 3 and for particular m, we are using 10 values of entropy as mentioned in Table 2. Table 3 shows the reduction of the 4,09,700 sample to small sample size.

Table 1. Frame size

Size of frame (N)	No. of frame per each time-series
173	4097/173 = 23
256	4097/256 = 16
512	4097/512 = 8
1024	4097/1024 = 4
2048	4097/2048 = 2

Table 2. Number of entropy value per each time-series

Size of frame (N)	No. of frame per each time-series	No. of entropy values per each time-series
173	23	30 * 23 = 690
256	16	30 * 16 = 480
512	8	30 * 8 = 240
1024	4	30 * 4 = 120
2048	2	30 * 2 = 60

Table 3. Reduction of sample size

Size of frame (N)	Final data after apply Apen (4097 * 100 = 4,09,700)
173	690 * 100 = 69000
256	480 * 100 = 48000
512	240 * 100 = 24000
1024	120 * 100 = 12000
2048	60 * 100 = 6000

2.2 Support Vector Machine (SVM)

We map input patterns into a higher dimensional feature space through using SVM (Support Vector Machine). In this high dimensional feature space, linear decision surface constructed. So, SVM is a linear classifier in the parameter space [15].

Let we take m dimensional training data set $xi = (1,\ldots, M)$ and their class labels be yi, where $yi = 1$ and $yi = -1$ for positive and negative classes respectively. In particular input space, linear separable data then the following decision function can be determined as,

$$\boldsymbol{D(x) = w^t g(x) + b} \tag{4}$$

Maps x into the l-dimensional space, we use $g(x)$ is a mapping function. B is a scaler and w is the vector in 1-dimensional space. If we separate data linearly, the decision function satisfies the following condition given below:

$$\begin{gathered} \boldsymbol{Yi = (w^t g(xi) + b) > = 1} \\ \text{Where, } i = 1,\ldots, M \end{gathered} \tag{5}$$

For an infinite number of decision functions I is linearly separable in the feature space then it satisfy Eq. (5). So, we require that the hyper-plane that have the largest margin between positive and negative class. The $D(x)/\|w\|$ is margin that contain minimum distance from the separating hyper-plane to the input data.

Assume that the margin is ρ, the following condition is to be satisfied

$$\frac{\boldsymbol{YiD(xi)}}{\|\boldsymbol{w}\|} \geq \boldsymbol{\rho} \quad \text{Where}, i = 1, \ldots, M \tag{6}$$

The product of ρ and $\|w\|$ is fixed

$$\boldsymbol{\rho}\|\boldsymbol{w}\| = \mathbf{1} \tag{7}$$

In order to obtain the optimal separating hyper-plane with contain maximum margin, w with the minimum $\|w\|$ that satisfying Eq. (6) found. From Eq. (7), this equations are solving this optimization problem. Minimizing Yi,

$$\boldsymbol{Yi} = (\boldsymbol{w}^t \boldsymbol{g}(\boldsymbol{xi}) + \boldsymbol{b}) > = \mathbf{1} \tag{8}$$

We introduce slack variable ξ, When training data are not linearly separable into Eq. (5) as follows subject to the constraints:

$$\boldsymbol{Yi} = (\boldsymbol{w}^t \boldsymbol{g}(\boldsymbol{xi}) + \boldsymbol{b}) > = \mathbf{1} - \boldsymbol{\xi i} \quad \xi i \geq 0 \text{ for } i = 1, \ldots, M \tag{9}$$

The optimal separating hyper-plane is determined so that the maximization of the margin and the minimization of the training error achieved. Minimizing

$$\frac{\mathbf{1}}{\mathbf{2}} \boldsymbol{w}^t \boldsymbol{w} + \frac{\boldsymbol{c}}{\mathbf{2}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{\xi}_{\boldsymbol{i}}^{\boldsymbol{p}} \tag{10}$$

Subject to the constraints:

$$\boldsymbol{Yi} = (\boldsymbol{w}^t \boldsymbol{g}(\boldsymbol{xi}) + \boldsymbol{b}) > = \mathbf{1} - \boldsymbol{\xi i} \quad \xi i \geq 0 \text{ for } i = 1, \ldots, M \tag{11}$$

Where C is a parameter that determines the trade-off between the maximum margin and the minimum classification error and p is 1 or 2. When $p = 1$, the SVM is called L1 soft margin SVM (L1-SVM), and when $p = 2$, L2 soft margin SVM (L2-SVM). In the conventional SVM, optimal separating hyper-plane obtained by solving the above quadratic programming problem. In this empirically and optimal results achieved using Radial Basis Function (RBF).

In first experiment, all 100 time-series of N and S is taken for training and testing. For frame size 173, entropy values are 690 for each time-series, so if we take 100 time-series, entropy values would be 69000 for one class and it is double (138000) by considering both seizure and non-seizure class. These procedures followed for all four features. Entropy values of both classes S and N for training and testing dataset for all frames is shown in Table 4 (Fig. 2).

Table 4. Number of entropy value for testing

Sr. no.	Time-series of N and S	Frame size	No. of entropy values for training	No. of entropy values for testing
1.	200	173	1,38,000	1,38,000
2.	200	256	96,000	96,000
3.	200	512	48,000	48,000
4.	200	1024	24,000	24,000
5.	200	2048	12,000	12,000

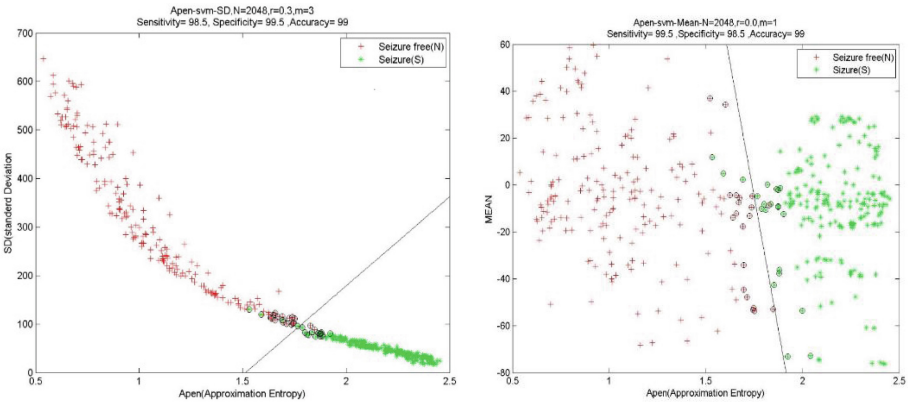


Fig. 2. ApEn for N and S file set for (a) $N = 2048$, $m = 3$, $r = 0.3$ with SD, (b) $N = 2048$, $m = 1$, $r = 0.0$ with mean

For frame size $N = 2048$ and $m = 3$ and $r = 0.3$, gets optimum accuracy for ApEn. From that, we get highest accuracy 99.00% of the feature ApEn with SD for experiment and frame size $N = 2048$ and $m = 1$ and $r = 0.3$, gets optimum accuracy for ApEn. From that, we get highest accuracy 99.00% of the feature ApEn with Mean for experiment.

For training purpose, all 50-time-series data for N and S taken and 100 time-series data, taken for testing. Entropy values of both classes S and N for training and testing dataset for all frames as shown in Table 5.

Table 5. Entropy value after 50% training

Sr no.	Frame size	Time-series of N and S for training	Time-series of N and S for testing	No. of entropy values for training
1.	173	100	200	69,000
2.	256	100	200	48,000
3.	512	100	200	24,000
4.	1024	100	200	12,000
5.	2048	100	200	6000

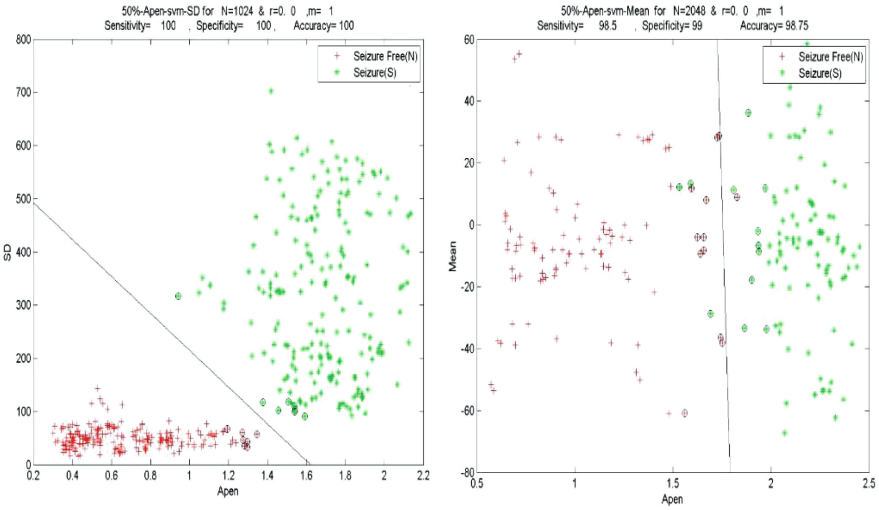


Fig. 3. ApEn for N and S file set after 50% training and testing (a) $N = 1024$, $m = 1$, $r = 0.0$ with SD, (b) $N = 2048$, $m = 1$, $r = 0.0$ with mean

The above Fig. 3 are for all optimum results of experiment feature dataset as shown in above Table 5. The figure shows the SVM classification for the seizure and normal class using radial basis kernel function. Where seizure is denoted by * and normal by +. The line is describing linear classification of the dataset. The o describes wrongly classify data points of opposite class.

2.3 Performance Parameters

2.3.1 Standard Deviation

Quantify the amount of variation or dispersion of a set of data values by using standard deviation. The standard deviation of a random variable like,

(1) Statistical population,
(2) Data set, or probability distribution is the square root of its variance [15].

2.3.2 Mean

The Mean is also called as a arithmetic mean of a sample. It is usually denoted by x. The x is the sum of the signals sampled values divided by the number of items in the sample [15].

2.3.3 Sensitivity

$$\text{Sensitivity} = \frac{\text{No. of true positive detected data points}}{\text{total no. of positive data points}} \tag{12}$$

Sensitivity considered for detection of seizure data [16].

2.3.4 Specificity

$$\text{Specificity} = \frac{\text{No. of true negative detected data points}}{\text{total no. of negative data points}} \tag{13}$$

Specificity considered for detection of non-seizure data [16].

2.3.5 Accuracy

$$\text{Accuracy} = \frac{(\text{TP}) + (\text{TN})}{\text{total no. of data points}} \tag{14}$$

TP = No. of true positive detected data points
TN = No. of true negative detected data points [16].

3 Experimentation Results

In our work, we have extracted the features from the EEG signal and classification done using SVM classifier in to two class seizure-free and seizure patient data. ApEn values is measure in form of m, r, and N. The values of m, r, and N are as follows:

(1) Number of Samples (m) = 1, 2, 3;
(2) Normalization Ratio (r) = 0%–90% of SD of the data sequence in increments of 10%;
(3) Frame Size (N) = 173, 256, 512, 1024 and 2048.

Approximation Entropy is extracted along with SD and mean. The randomness of EEG signal were extracted in the features, based on different size of frame (N), number of samples values (m) and normalized ratio (k). From the set of features, ApEn with SD and mean, are used for classification using the SVM classifier.

We have used BONN dataset for EEG signals which is publicly available online and described in Andrzejak et al. [17]. The EEG dataset contains both seizures and non-seizures. The Bonn dataset consists five subsets (Z, O, N, F, and S) each containing 100 single-channel EEG signals, each signal of 23.6 s in duration with the sampling rate of 173.61 Hz.

EEG recordings of five healthy volunteers with eyes open (Z) and closed (O) have been recorded on the surface, using standard electrode placement scheme. The signal F and S are seizure free subset. These two are recorded in seizure-free intervals from five patients in the epileptogenic zone (F-Seizure free) and from the hippocampal formation

of the opposite hemisphere of the brain (N-seizure free). The set S is contained seizures signal which gives an ictal activity by using with the same 128-channel amplifier system with an average common reference all EEG signals are recorded. In the proposed work classification of the N (Seizure free class) and S (Seizure class) is done by using ApEn (Approximate Entropy) feature extraction and reduction and SVM as

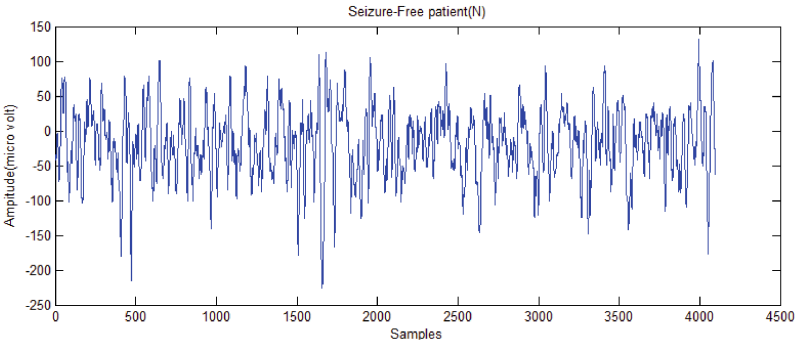


Fig. 4. N (Seizure-free or Normal EEG) Patient Class

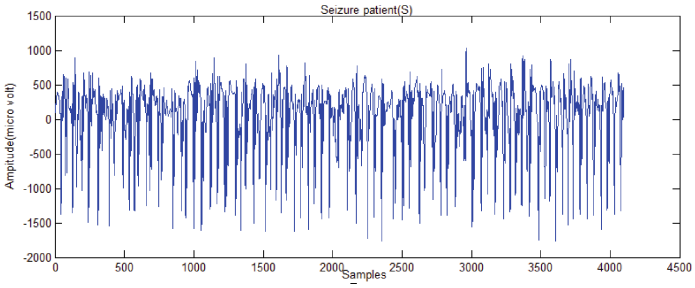


Fig. 5. S (Seizure or Epileptic EEG) Patient Class

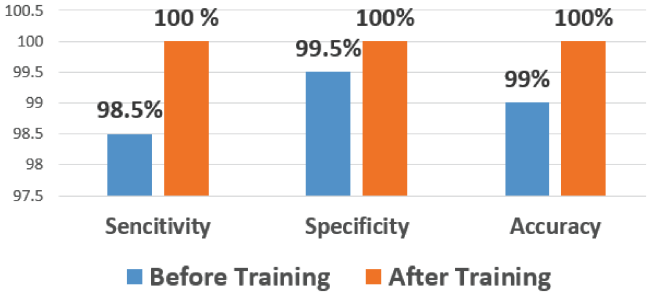


Fig. 6. Classification accuracy, sensitivity and specificity at before training and after 50% training for N (Seizure free) and S (Seizure Class)

classifier. Figures 4 and 5 show N (Seizure free-patient) and S (Seizure-patient) EEG signals, respectively. It is containing only seizure impulse. Here each dataset contain 100 time-series. Each signal contains 4097 samples (Fig. 6).

For all 100 EEG data sets, 50 data sets are used for training and the others are used for testing using SVM classifier. SVM classifier is used to classify unknown data properly. The highest accuracy is 100% for the feature set ApEn with SD for frame size $N = 1024$, sample value $m = 1$ and normalization ratio $r = 0.0$. In the proposed method accuracy achieved up to 100% for the feature set ApEn with SD. For training and testing purpose we get different accuracy, sensitivity and specificity as shown in graph.

Table 6. Comparison of methodology for same dataset

Methodology	Subset	Classification accuracy (%)
Permutation entropy (PE) and SVM classifier [14]	N and S	88.83
Empirical mode decomposition (EMD) [16]	N and S	95.33
Clustering and SVM classifier [18]	S and N	97.69
SODP and artificial neural network (ANN) classifier [15]	S and N	97.75
Empirical mode decomposition (EMD) and phase space representation (PSR) [19]	N and S	98.67

As shown in Table 6 all the papers are worked on Bonn dataset and they achieved maximum accuracy is 98.67%. In the proposed method accuracy achieved up to 100% for the feature set ApEn with SD.

4 Conclusion

We have extracted the features from the EEG signal and classification done using SVM classifier in to two class seizure and normal. Approximation entropy is extracted along with SD and mean. The randomness of EEG signal were extracted in the features, based on different size of frame (N), no. of samples values (m) and normalized ratio (r). From the set of features ApEn with SD, ApEn with mean, were used for classification using the SVM classifier. The highest classification accuracy is 100% for N and S class.

References

1. Lehnertz, K., Mormann, F., Kreuz, T., Andrzejak, R.G., Rieke, C., David, P., Elger, C.E.: Seizure prediction by nonlinear EEG analysis. IEE Eng. Med. Biol. Mag. **22**(1), 57–63 (2003). Article in IEEE engineering in medicine and biology magazine, Research gate January 2003
2. Mc Grogan, N.: Neural network detection of epileptic seizures in the electroencephalogram (1999). http://www.newox.ac.uk/˜nmcgroga/work/transfer

3. Srinivasan, V.: Approximate entropy-based epileptic EEG detection using artificial neural networks. IEEE Trans. Inf. Technol. Biomed. **11**(3), 288–295 (2007)
4. Güler, N.F., Übeyli, E.D., Güler, I.: Recurrent neural networks employing Lyapunov exponents for EEG signal classification. Exp. Syst. Appl. **29**(3), 506–514 (2005)
5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The physiology of the grid: an open grid services architecture for distributed systems integration. Technical report, Global Grid Forum (2002)
6. Übeyli, E.D.: Lyapunov exponents/probabilistic neural networks for analysis of EEG signals. Exp. Syst. Appl. **37**(2), 985–992 (2010)
7. Accardo, A., Affinito, M., Carrozzi, M., Bouquet, F.: Use of the fractal dimension for the analysis of electroencephalographic time series. Biol. Cybern. **77**(5), 339–350 (1997)
8. Liang, S.F., Wang, H.C., Chang, W.L.: Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. EURASIP J. Adv. Sig. Process. **2010** (2010). Article ID 853434
9. Pachori, R.B.: Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition. Res. Lett. Sig. Process. **2008** (2008). Article ID 293056
10. Oweis, R.J., Abdulhay, E.W.: Seizure classification in EEG signals utilizing Hilbert-Huang transform. BioMed. Eng. OnLine **10**, 38 (2011)
11. Pachori, R.B., Bajaj, V.: Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition. Comput. Meth. Programs Biomed. **104**(3), 373–381 (2011)
12. Bajaj, V., Pachori, R.B.: EEG signal classification using empirical mode decomposition and support vector machine. In: Proceedings of the International Conference on Soft Computing for Problem Solving, AISC 131, 20–22 December 2011, Roorkee, India, pp. 623–635 (2011)
13. Li, S., et al.: Feature extraction and recognition of ictal EEG using EMD and SVM. Comput. Biol. Med. **43**(7), 807–816 (2013)
14. Nicolaou, N., Georgiou, J.: Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. Elsevier Trans. Exp. Syst. Appl. **39**, 202–209 (2012)
15. Pachori, R.B., Patidar, S.: Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode function. Elsevier Trans. Comput. Meth. Program. Biomed. **113**, 494–502 (2014)
16. Bajaj, V., Pachori, R.B.: Classification of seizure and non seizure EEG signals using empirical mode decomposition. IEEE Trans. Inf Technol. Biomed. **16**(6), 1135–1142 (2012)
17. Andrzejak, R.G., et al.: Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64** (2001). Article ID 061907
18. Siulya, Li, Y., Wen, P.: Clustering technique-based least square support vector machine for EEG signal classification. Elsevier Trans. Comput. Meth. Program. Biomedic. **104**, 358–372 (2011)
19. Sharma, R., Pachori, R.B.: Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. Elsevier Trans. Exp. Syst. Appl. **12**, 1106–1117 (2015)

Comparative Analysis of PSF Estimation Based on Hough Transform and Radon Transform

Mayana Shah[1(✉)] and Upena Dalal[2]

[1] CKPCET, Surat, India
mayna.shah@ckpcet.ac.in
[2] SVNIT, Surat, India
udd@eced.svnit.ac.in

Abstract. Blind image motion deblurring (BID) is in great demand to recover the original image from its degraded observation. Motion blur is the effect of relative movement between camera and object during shutter opening. Restoring the information requires estimation of Point spread function (PSF) and use this PSF for deblurring task. PSF estimation plays important role in motion deblurring and mis-specification of kernel can lead to structural distortion in deblurred image. In this paper, we have proposed the comparative analysis of PSF estimation methods in modified cepstrum domain based on Hough transform and Radon transform. Experimentation is done on standard image and estimated parameters are compared for motion blur of different length and degrees. Conclusions are drawn on the basis of simulation study on Matlab for standard image.

Keywords: Image deblurring · Image restoration · Motion blur
PSF estimation

1 Introduction

The BID is mainly useful in almost all imaging related applications as normally there is always a chance of camera shake during the photo capturing process. The Deblurring result accuracy depends on the accuracy of PSF estimation. Once the PSF is accurately estimated, non-blind deblurring is used to get restored image. In most of the research work blurring process can be modelled by convolution formula as [1]:

$$g(x, y) = h(x, y) * f(x, y) \tag{1}$$

Equation (1) show that $g(x, y)$ captured degraded image is nothing but convolution of pristine image f with degradation function h. Here (x, y) Indicates spatial coordinates and "*" is the convolution operation. There are diverse techniques for image deblurring with simultaneous or separate PSF estimation. Comprehensive overview of all techniques is given by Wang and Tao in [2]. Comparison of all such techniques are given in Table 1.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 86–96, 2018.
https://doi.org/10.1007/978-3-319-73712-6_9

Table 1. Comparison of image deblurring techniques with simultaneous or separate PSF estimation

Type	Functionality	Advantage and limitations
Statistical methods	MAP Variational Edge prediction	Convergence problem Produce good results if converge to right solution Slow Requires prior information Face problem in restoring image with multiple neighboring edges [3–7]
Regularization	Tihkhonov Total Variation Dictionary learning	Regularization parameter setting effect the solution, Trade-off between performance and complexity [8–10]
Parametric methods	Spectral Cepstrum Modified Cepstrum	Simpler approach - less computation [11–16, 23, 24]
Hardware based methods	Use of gyroscopes Coded shutter Coded aperture	Corrects the blur before information is recorded on the sensor Costly High quality [17–20]
Multi-channel image restoration	Dual Camera	For multi images we require precise registration Slow computation [21, 22]

Our method is a parametric approach which uses a mathematical model of a uniform motion blur kernel h expressed in terms of parameters length L and theta θ as [1]:

$$h(x, y, L, \theta) = \begin{cases} \frac{1}{L}, & if\ \sqrt{x^2 + y^2} \leq L/2, x/y = -\tan^{-1}\theta \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

To remove the blurring effect the parameters length L and theta θ should be decided accurately. Spectral representation of a blur kernel h is a sinc function as motion blur is kind of rectangular function. Zeros of the Sinc function helps to find out length L and orientation of sinc is in perpendicular direction of motion. Existing methods use log-spectral representation [11, 12] and Cepstral domain representations [13–15] of the blurred input image to obtain blur kernel estimation, but it may lead to erroneous angle estimation because of no of parallel stripes in magnitude spectrum. So to thin out central lobe we use dual operated log spectrum termed as modified cepstrum defined as follows [16]:

$$C = \log\{|\mathscr{F}\{\log\{|\mathscr{F}(g(x, y))|\}\}|\} \tag{3}$$

Despande shown modified cepstrum [16] has potential clues to identify PSF resulted from thin line segment view of central thick stripe in spectrum of blur image. To find out the direction of line segment there are two approaches first one by Hough

transform and second one by Radon transform. In this paper, we have presented a comparison of both approaches and shown that Radon transform based technique produces superior results compare to Hough transform based method. Hough transform is flexible to find lines in images and one can easily represent broken lines as a joint line but at specific angles its performance degrades and accuracy decreases.

Comparison of these two major approaches over wide range of blur parameter variation needed to be explored and effort of such comparative analysis is done in paper. The rest of the paper is organized as follows: Sect. 2 describes Hough Transform based parameter estimation and results. Radon Transform based parameter estimation and results is presented in Sect. 3. The comparison is discussed in Sect. 4 and the conclusions are summed up in Sect. 5.

2 Hough Transform Based Parameter Estimation

Modified cepstrum represents thin line segment in direction of motion blur. To find out the direction of line segment there are two approaches first one by Hough transform and second one by Radon transform. In our previous work [23, 24] Hough transform based algorithm were discussed that has been summarized in Algorithm 1.

Algorithm 1

1. Obtain modified cepstrum from grayscale blurred image (Fig.1(d))
2. Extract fourth bit plane of modified cepstrum image and use canny edge detection for finding the line segment. (Fig.1(e) and Fig.1(f))
3. Apply Hough transform on step 2 image and find Motion blur angle is identified as 180- the peak in Hough transform
4. Step one image is grayscale transform by threshold value 0.6. and rotate it by negative value of estimated angle (Fig.1(g))
5. Step 5 image is converted to 1-D by averaging of columns. Twin peak pattern is obtained with center peak and two side peaks. Averaging distance from center peak to two side peak is considered as blur length. (Fig.1(h))

Results for cameraman image of size 256×256 is presented in Table 2 for degradation by different lengths ($5 \leq L \leq 80$) and different directions ($10 \leq \theta < 170$). Error equals to the estimated value minus the real value. The results are as shown in Fig. 1.

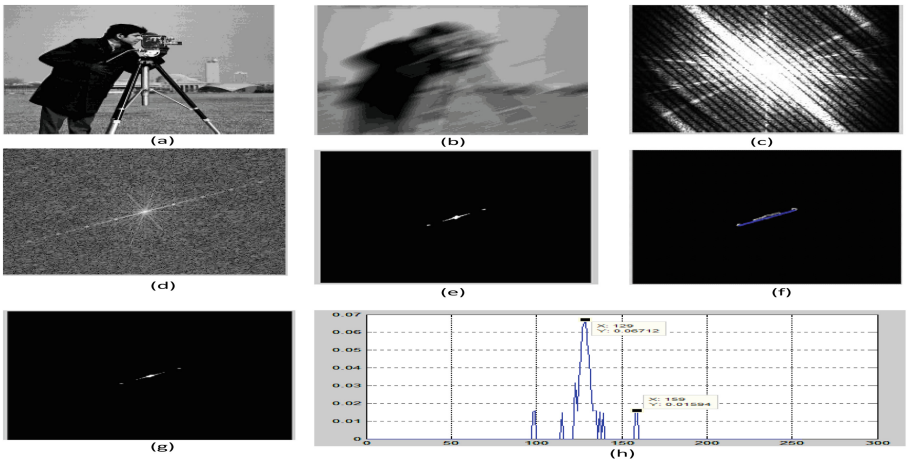


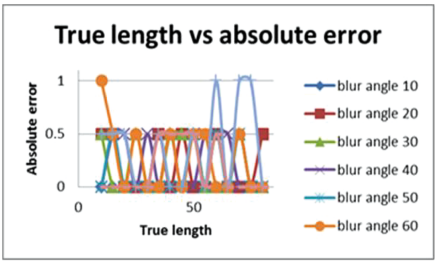
Fig. 1. Results of proposed algorithm for uniform blur without noise (a) original cameraman (b) blurred image with L = 30 pixels and θ = 30°, (c) log spectrum (d) modified cepstrum domain (e) thin line segment extracted from fourth bitplane of modified cepstrum (f) edge detection and Hough based angle estimation θ^ = 28° (g) grayscale transformation of modified cepstrum with threshold 0.6 (h) twin peak representation and blur length estimation L^ = 30

Based on Estimated parameters in Table 2 graph is plotted for true length Vs absolute error for various blur direction and result is shown in Fig. 2(a) and (b). From Fig. 2(a) and (b) it is observed that maximum error in blur length estimation is 1 pixel.

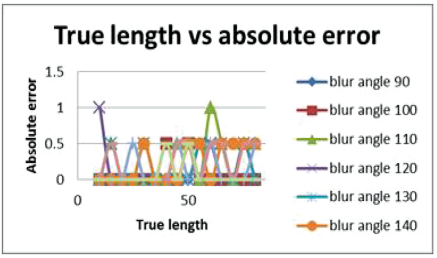
A result for Blur angle estimation for camera man image is given in Table 3. Based on Estimated parameters graph is plotted for true angle Vs absolute error for various blur length and result is shown in Fig. 3(a) and (b). Graphical representation shows that maximum error in blur angle estimation is 4°.

Table 2. Results for blur length estimation for cameraman image.

True L in *pixels*		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
		Estimated blur length															
Blur angle in degrees	10	5	10	15	20	25	30	35.5	40	45.5	50	55	60	65	70	75	80
	20	5.5	9.5	15	20	25	31	35	39.5	45	50	54.5	60	65	70	75.5	80
	30	5.5	10	15	19.5	25	30	35.5	40.5	45.5	50	55	60	65.5	70	75	80
	40	5	10.5	14.5	20	24.5	30	35	40	45	50.5	55.5	60.5	65	70	75	80
	50	5	10.5	15	20	25	30	35	40	45.5	50	55.5	60	64.5	70	75	80
	60	6	10.5	15	19.5	25	30	35.5	40	45.5	49.5	55	60	64.5	70	75	79.5
	70	5.5	9.5	15.5	20	25	30.5	35	40	45.5	50	54	60	66	69	75	80
	80	5	10	15	20	25	29.5	35.5	39.5	45.5	50	55.5	60	65	70	75	80
	90	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	100	5	10	15	20	25	30	35.5	40	45.5	50	55	60	65	70	75	80
	110	5	9.5	15	20	25	30	35	39.5	45.5	50	54	60.5	65	69.5	75.5	79.5
	120	6	10	15	20	25	30	35	40	45	50.5	54.5	60	65	70.5	75	80
	130	5	10.5	15	20	25.5	30	35	40	45	50.5	54.5	60.5	65	69.5	75	80
	140	5	10	15	20	25.5	30	35	40	45.5	50.5	55	60.5	64.5	69.5	75.5	80
	150	5	10	15	19.5	25	30	35	40.5	45	50	54.5	60	65	70	74.5	79.5
	160	5	9.5	15	20	25	30	35	39.5	45.5	50	54.5	60.5	65	69.5	75.5	80
	170	5	10	15	20	25	30	35.5	40	45.5	50	55	60	65	70	75	80



(a)

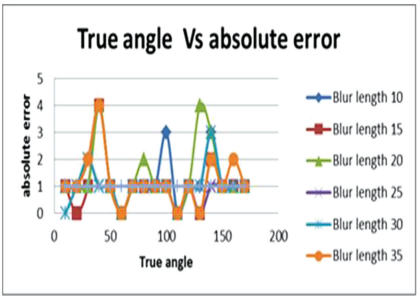


(b)

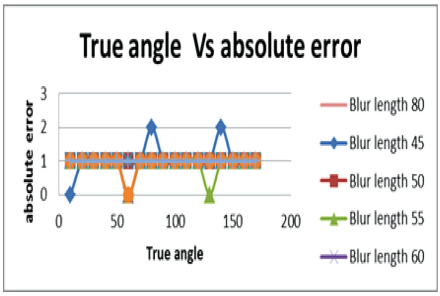
Fig. 2. (a) True length vs absolute error for blur angle variation 10° to 60° (b) True length vs absolute error for blur angle variation 90° to 140°

Table 3. Results for blur angle estimation for cameraman image.

L in *pixels*		10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
		Estimated blur angle in degrees														
Blur angle in degrees	10	9	9	9	9	10	9	9	10	9	9	9	9	9	9	9
	20	19	20	19	19	19	19	19	19	19	19	19	19	19	19	19
	30	28	29	29	28	28	28	29	29	29	31	29	29	29	29	31
	40	44	44	44	39	39	44	39	39	39	39	39	39	39	39	39
	50	49	49	49	49	49	49	49	49	49	49	49	49	49	49	49
	60	60	60	60	60	60	60	59	59	59	60	60	59	60	59	60
	70	71	69	69	69	69	69	69	69	69	69	69	69	69	69	69
	80	79	79	78	79	79	79	79	78	79	79	79	79	79	79	79
	90	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89
	100	97	99	99	99	99	99	99	99	99	99	99	99	99	99	99
	110	110	110	110	110	110	110	109	109	109	109	109	109	109	109	109
	120	119	119	119	119	119	119	119	119	119	119	119	119	119	119	119
	130	131	130	134	130	129	130	129	129	129	130	129	129	129	129	129
	140	137	138	137	139	137	138	139	138	139	139	139	139	139	139	139
	150	149	149	149	149	149	149	149	149	149	149	149	149	149	149	149
	160	159	159	159	159	159	158	159	159	159	159	159	159	159	159	159
	170	169	169	169	169	169	169	169	169	169	169	169	169	169	169	169



(a)



(b)

Fig. 3. (a) True angle vs absolute error for blur length 10 to 35 pixels (b) True angle vs absolute error for blur length 45 to 80 pixels

3 Radon Transform Based Parameter Estimation in Modified Cepstrum Domain

The Proposed scheme uses Radon Transform for blur direction identification. Radon Transform is kind of integral transform that computes the integration of a function along straight lines and useful to detect linear representations in images. Line in the modified cepstrum image will be represented by a peak in the Radon transform whose location determines the parameters of the line. Radon transform along this direction usually has larger variations so peak in radon transform corresponds to max value of variance. As the integration is along the perpendicular direction to a line, principal direction of motion blur is obtained by subtracting 90° from the max value of Radon transform. The detail method is discussed in Algorithm 2 and result is shown in Fig. 4.

Algoritham 2

1. Obtain modified cepstrum from grayscale blurred image (Fig.4(d))
2. Perform gray scale transformation on step 1 image with threshold=0.6 (Fig.4(e))
3. Find the principal direction using Radon transform as peak in Radon transform-90° (Fig.4(f)).
4. Rotate the grayscale transformed modified Cepstrum image of step 2 by negative value of estimated angle. (Fig.4(g))
5. Convert the 2-D matrix of step 4 to 1-D by taking the averages of columns. It will show a twin peak pattern. (Fig.4(h))
6. The distance between the central peak and first larger peak on either side is nothing but the estimated blur length in pixels (Fig.4(h)).

PSF estimation Algorithm 2 discussed is applied to Lena Image which was degraded by different lengths ($10 \leq L \leq 80$) and different directions ($10 \leq \theta° < 170$). The results are presented in Tables 4 and 5. Error equals to the estimated value minus the real value. The results are as shown in Fig. 4.

Based on Estimated parameters in Table 4 graph is plotted for true length Vs absolute error for various blur direction and result is shown in Fig. 5(a) and (b). It is observed that maximum error in blur length estimation is 1 pixel. A result for Blur angle estimation for Lena image of size 256×256 is given in Table 5. Based on Estimated parameters graph is plotted for true angle Vs absolute error for various blur length and result is shown in Fig. 6(a) and (b). Graphical representation shows that maximum error in blur angle estimation is 1°.

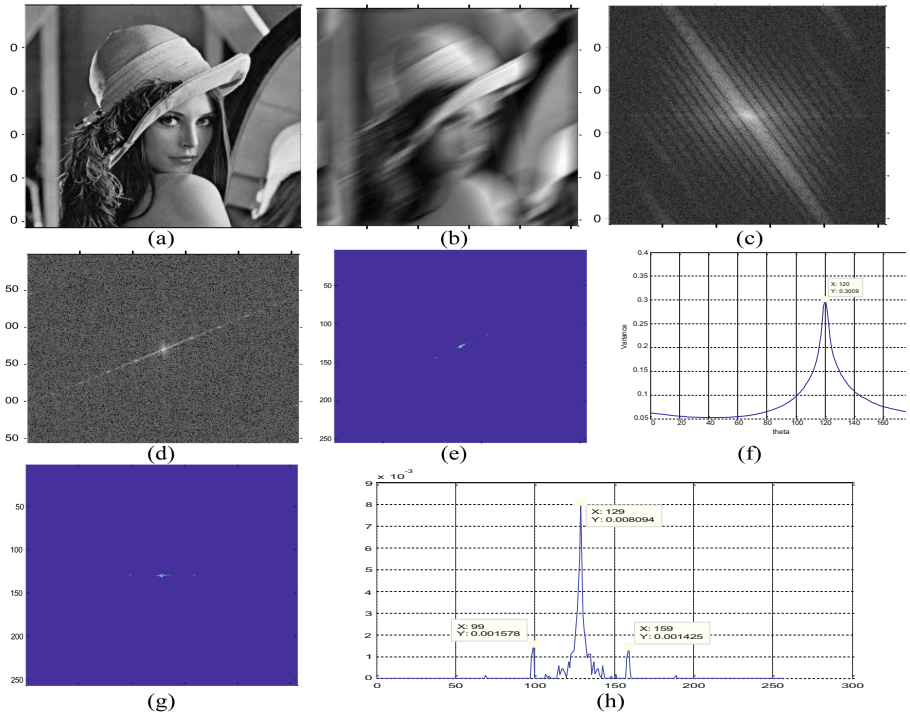


Fig. 4. Results of proposed algorithm for uniform blur (a) original Lena image (b) blurred image with L = 30 pixels and θ = 30°, (c) log spectrum (d) modified cepstrum domain (e) thin line segment extracted from Gray scale transform of modified cepstrum with threshold 0.6 (f) Radon transform based blur direction estimation θ^ = 30° (g) anticlockwise rotation of grayscale transformed modified cepstrum (h) twin peak representation and blur length L^ = 30 pixels

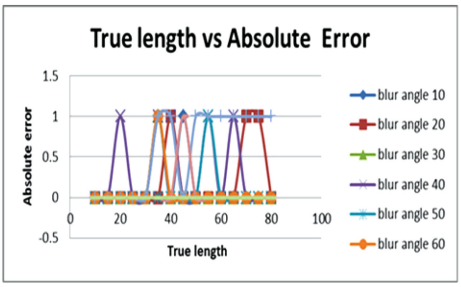
4 Comparison

Following observations can be made from experiments of Sects. 2 and 3:

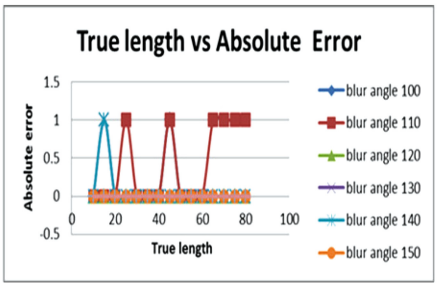
- In case of Hough transform based blur angle estimation at specific angles such as 40° and 140° accuracy reduces and it shows maximum error of 4° and there are more chance of error for almost whole range of blur length.
- In case of Radon transform based blur length estimation accuracy is 87.45%. It shows great impact on deblurred image as deblurred image output quality is highly dependent on accuracy of PSF estimation in particular blur length.
- In case of Radon transform based blur angle estimation accuracy is 95.68% and most chance of error in blur angle measurement is in case where the blur length is lower or equal to ten. In case of Radon transform based blur angle estimation maximum error reduces to 1°.
- Almost zero error in blur angle estimation after blur length 15 pixels in Radon transform based approach.

Table 4. Results for blur length estimation for LENA 256 × 256 image

True L *in Pixels*		10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
		Estimated blur length														
Blur angle 0°	10	10	15	20	25	30	36	40	46	50	55	60	65	70	75	80
	20	10	15	20	25	30	35	41	45	50	55	60	65	71	76	80
	30	10	15	20	25	30	36	40	45	50	55	60	65	70	75	80
	40	10	15	19	25	30	36	41	45	50	56	60	66	70	75	80
	50	10	15	20	25	30	35	40	45	50	56	60	65	70	75	80
	60	10	15	20	25	30	36	40	45	50	55	60	65	70	75	80
	70	10	15	20	25	30	34	41	45	49	54	59	64	71	74	81
	80	10	15	20	25	30	35	40	46	50	55	60	65	70	75	80
	90	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	100	10	15	20	25	30	35	40	46	50	55	60	65	70	75	80
	110	10	15	20	26	30	35	40	46	50	55	60	64	69	76	81
	120	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	130	10	14	20	25	30	35	40	45	50	55	60	65	70	75	80
	140	10	14	20	25	30	35	40	45	50	55	60	65	70	75	80
	150	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	160	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	170	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80



(a)

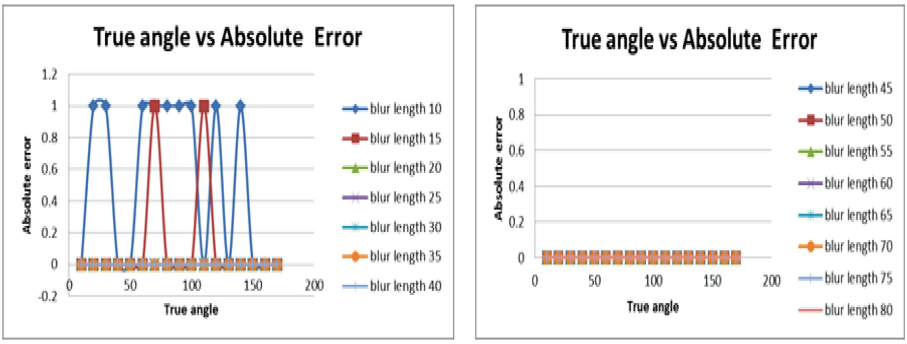


(b)

Fig. 5. (a) True length vs absolute error for blur angle 10° to 60° (b) True length vs absolute error for blur angle 100° to 150°

Table 5. Results for blur angle estimation for LENA 256 × 256 image.

L *in Pixels*		10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
		Estimated blur angle														
Blur angle θ°	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	20	21	20	20	20	20	20	20	20	20	20	20	20	20	20	20
	30	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30
	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
	60	61	60	60	60	60	60	60	60	60	60	60	60	60	60	60
	70	71	71	70	70	70	70	70	70	70	70	70	70	70	70	70
	80	79	80	80	80	80	80	80	80	80	80	80	80	80	80	80
	90	89	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	100	101	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	110	110	109	110	110	110	110	110	110	110	110	110	110	110	110	110
	120	119	120	120	120	120	120	120	120	120	120	120	120	120	120	120
	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130
	140	139	140	140	140	140	140	140	140	140	140	140	140	140	140	140
	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150
	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
	170	170	170	170	170	170	170	170	170	170	170	170	170	170	170	170



(a) (b)

Fig. 6. (a) True angle vs absolute error for blur length 10 to 40 pixels (b) True angle vs absolute error for blur length 45 to 80 pixels

5 Conclusion

Comparative analysis of PSF parameters estimation is shown in the paper concentrate on vital parameters like wide variation of blur extent and increasing accuracy of PSF estimation. The discussed PSF estimation is done in modified cepstrum domain using Hough transform and radon transform. Comparative analysis shows that even though Hough transform is flexible to use in case of broken line segments its PSF estimation accuracy is lower compare to radon transform based approach in large blur parameter variation range. Experimental results show that radon transform based approach is more accurate for wider variation of blur extent.

References

1. Gonzalez, R., Woods, R., Eddins, S.: Digital Image Processing Using MATLAB. Pearson Prentice-Hall, Upper Saddle River (2004)
2. Wang, R., Tao, D.: Recent progress in image deblurring. arXiv preprint arXiv:1409.6838 (2014)
3. Fergus, R., et al.: Removing camera shake from a single photograph. ACM Trans. Graph. **25**(3), 787–794 (2006)
4. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: CVPR (2011)
5. Xu, L., Zheng, S., Jia, J.: Unnatural L0 sparse representation for natural image deblurring. In: CVPR (2013)
6. Cho, J., Lee, S.: Fast motion deblurring. ACM Trans. Graph. **28**(5), December 2009
7. Sun, L., Cho, S., Wang, J., Hays, J.: Edge-based blur kernel estimation using patch priors. In: ICCP (2013)
8. Groetsch, C.W.: The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Pitman, London (1984)
9. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. SIAM Multiscale Model. Sim. **4**, 460–489 (2005)
10. Elad, M.: Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer, Heidelberg (2010)
11. Gennery, D.B.: Determination of optical transfer function by inspection of frequency domain plot. J. Opt. Soc. Am. **63**(12), 1571–1577 (1973)
12. Lokhande, R., Arya, K.V., Gupta, P.: Identification of parameters and restoration of motion blur images. In: ACM Symposium on Applied Computing, pp. 301–305 (2006)
13. Cannon, P.: Blind deconvolution of spatially invariant image blurs with phase. IEEE Trans. Acoust. Speech Sign. Process. **24**(1), 56–63 (1976)
14. Chen, C.H., Rui, Z.: Image restoration for linear local motion blur based on cepsturm. In: International Conference on Genetic and Evolutionary Computing (2012)
15. Park, J., Kim, M., Chang, S., Lee, K.H.: Estimation of motion blur parameters using cepstrum analysis. In: IEEE 15th International Symposium on Consumer Electronics (ISCE), pp. 406–409 (2011)
16. Deshpande, A.M., Patnaik, S.: A novel modified cepstral based technique for blind estimation of motion blur. Int. J. Light Electron. Opt. **125**(2), 606–661 (2014)

17. Agrawal, A., Xu, Y.: Coded exposure deblurring: optimized codes for PSF estimation and invertibility. In: IEEE Conference on Computer Vision and Pattern Recognition, New York, pp. 2066–2073 (2009)
18. Zhou, Y., Lin, S., Nayar, S.K.: Coded aperture pairs for depth from defocus and defocus deblurring. Int. J. Comput. Vis. **93**(1), 53–72 (2011)
19. Levin, A., et al.: Image and depth from a conventional camera with a coded aperture. ACM Trans. Graph. **26**(3) (2007)
20. Veeraraghavan, A., et al.: Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans. Graph. **26**(3), July 2007
21. Hiura, S., Matsuyama, T.: Depth measurement by the multi-focus camera. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 953–959 (1998)
22. Ben-Ezra, M., Nayar, S.K.: Motion-based motion deblurring. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 689–698 (2004)
23. Shah, M.J., Dalal, U.: Blind estimation of motion blur kernel parameters using cepstral domain and Hough transform. In: Fifth International Conference on Advances in Computing, Communication and Informatics – ICACCI (2014)
24. Shah, M.J., Dalal, U.: Hough transform and cepstrum based estimation of spatial-invariant and variant motion blur parameters. In: International Conference on Advances in Electronics, Computers and Communications (ICAECC) (2014)

Compressive Sensing Based Image Reconstruction

Sherin C. Abraham[1], Ketki Pathak[2(✉)], and Jigna J. Patel[1]

[1] Electronics & Communication Department, Dr. S. & S. S. Ghandhy Government College, Surat, India
sherincheeranabraham@gmail.com, jigna2012me@gmail.com
[2] Sarvajanik College of Engineering and Technology, Surat, India
ketki.joshi@scet.ac.in

Abstract. Compressive Sensing is novel technique where reconstruction of an image can be done with less number of samples than conventional Nyquist theorem suggests. The signal will pass through sensing matrix wavelet transformation to make the signal sparser enough which is a criterion for compressive sensing. Different levels of wavelet decomposition are also analyzed in this paper. The performance further can be improved by using DARC prediction method. The prediction error signal transmitted through OFDM channel. The reconstructed image should be better in both PSNR and bandwidth. Medical field especially in MRI scanning, compressive sensing can be utilized for less scanning time.

Keywords: Compressive sensing · Wavelet transform · Sparsity
DARC prediction · Predictive coding · LZW encoder

1 Introduction

Compressive sensing (CS) is new compression technique where with fewer samples of measurements is enough to reconstruct the image with good visual quality. The samples required are much lesser than Nyquist criterion suggests, but reconstruction is more complex in CS whereas linear in conventional compression. Now CS is actively researched in applications like MRI, RADAR, single pixel camera, etc. [1].

We consider the application in medical field. MRI is slow process due to large number of data need to be collected while scanning a patient. With the help of CS we can reduce the number of samples or skip certain acquisitions, which will benefit patient with less radiation exposure since scan time reduction is exactly proportional to the degree of under-sampling [2].

In CS, there are three main principles – Sparsity, measurements taking and nonlinear reconstruction. The signal should be sparse – Information rate contained in the image should be much less than bandwidth - to undergo CS. If it's not sparse enough; we need to undergo the transformation of the image to make it sparse. We took wavelet transform as sparsity inducing matrix in this paper. The reconstruction of signals from lesser samples can only possible if the chosen sparsity matrix and measurement matrix follows Restricted Isometric Property. The incoherence between these matrices is necessary for this. There are two approaches for reconstructing image at receiver side – basis pursuit

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 97–105, 2018.
https://doi.org/10.1007/978-3-319-73712-6_10

and greedy algorithm. These nonlinear techniques will result in good quality reconstructed image [3]. In short, CS can able to reduce sampling and computation costs for sensing signals that have a sparse or compressible representation. The sampling or acquisition method is by multiplying the sparse signal with the measurement matrix as in Fig. 1.

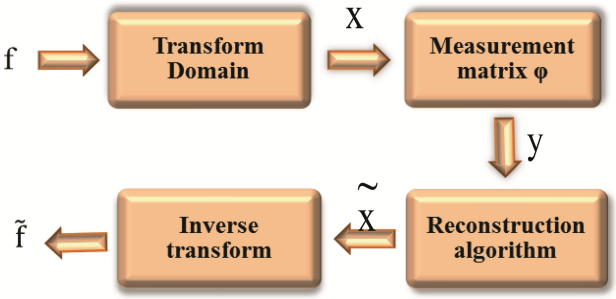


Fig. 1. Basic block diagram of compressive sensing

The channel considered as OFDM, because of its wide use in communication field. The OFDM technique is used for high data rate wireless communication due to high multipath interference rejection [4]. Through this paper, we aim to reconstruct the image with good visual quality which is required in medial field applications but with less bandwidth required for channel transmission. The importance of visual quality in medical field is clear in Fig. 3 and it can be measured through PSNR parameter.

In paper [5], the authors introduced a novel compressive sensing based prediction measurement (CSPM) encoder. The sparse image undergoes CS and these measured values pass to CSPM. In CSPM, the measured matrix undergoes linear prediction and LZW encoding. This CSPM encoder can achieve significant reduction in data storage and save transmission energy.

The wavelet transformed image has high frequency coefficients that are sparse and the low frequency coefficients that are not sparse. The low-frequency coefficients contain most energy of the image and have coherence nature. In paper [6], they measured (CS) the high-frequency sub band coefficients, and keep the low-frequency sub-band coefficients unchanged.

In this paper we have chosen deterministic matrix such as Hadamard matrix, Toeplitz matrix, random matrix such as Gaussian matrix, Bernoulli matrix as measurement matrices and to attain more sparsity, different levels of wavelets are used. The nonlinear reconstruction methods used are Orthogonal Matching Pursuit (OMP) and L1 minimisation technique.

2 Basic Block Diagram

The signal which is K-sparse in one domain can be reconstructed from another domain which has cK non adaptive linear projections where c is small constant. The sparsity matrix and measurement matrix should be incoherent for good reconstruction.

Let x is a real valued, finite length, one dimensional, discrete time signal which is an Nx1 vector in R^N. When x is K-sparse, i.e. K << N, the signal x can undergoes CS.

Thus signal x can be written as

$$X = \Psi f \tag{1}$$

where f is the representation of signal x in another domain Ψ.

Let y be the measured vector of size $M \times 1$ and $\Phi \in R^{MxN}$ be the measurement matrix.

$$y = \Phi\Psi f = Acsf = \Phi x \tag{2}$$

where $Acs = \Phi\Psi$ is the sensing matrix of $M \times N$ [7].

3 Proposed Block Diagram

In Fig. 2, we propose a new method based on CSPM encoder. First the image will undergo 2-D wavelet transform. Different levels of wavelet decomposition are used to make image sparser. In paper [5], the whole wavelet coefficients are undergoing CS together and measured using a single measurement matrix. Low frequency components contain coarse information and high frequency components contain detail information. The processing of high frequency components together with low frequency components will not exploit the coherent nature of LL band and this lead to performance degradation of the reconstructed image. So in this paper, CS will be applied separately for high frequency bands and low frequency bands. When applying CS to the sparse signal, we need to take a measurement which is inner product of the signal with fixed matrix. This matrix should follow the Restricted Isometry Property and incoherence. Gaussian matrix, Bernoulli matrix, Hadamard matrix and Toeplitz matrix are used as measurement matrices in this paper.

The prediction error calculated from the measured values which add more sparsity, which enhance the performance. Differential adaptive run coding (DARC) prediction method is used in this paper and it gives better results than linear prediction which is used in paper [5]. Before transmission we need to encode the values. This will helps in reducing the number of measurements, storage space and bandwidth. LZW encoding and decoding which has better speed in running dictionary operations is implemented in this paper. Then the encoded signal passes through OFDM channel. The encoded data are interleaved to avoid burst errors without change the throughput or data rate. In order to avoid the inter-symbol interference (ISI) due to multipath delay spread, a cyclic prefix (CP) is inserted in each OFDM symbol prior to transmission.

At receiver side, we need to decode the signal received and predict the values. This is passed for reconstruction of wavelet coefficients using any basic pursuit or greedy

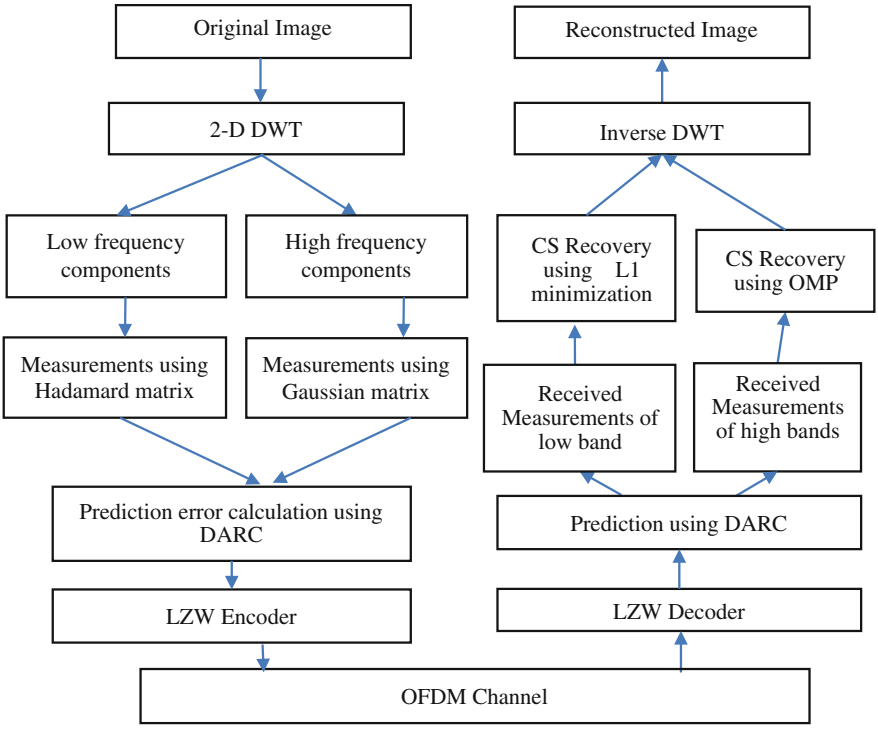


Fig. 2. Proposed block diagram

algorithms. The OMP and L1 minimization method is used for reconstruction in this paper. More iteration in OMP will give better results. After recovery, the image undergoes Inverse wavelet transform. The proposed diagram with techniques used in this paper is given in the Fig. 2.

4 Results

The image should be sparse to apply CS on it. After applying wavelet transform the signal will get sparser as shown in [7]. As we already discussed the LL band and high bands are processed separately. The high bands are applied with Gaussian measurement matrix since its random nature and LL band applied with different measurement matrix. The prediction error is calculated and the values are encoded by LZW algorithm. This encoded data passed through OFDM channel. For CS recovery we used L1 minimisation technique for LL band and OMP for high bands (Table 1 and Fig. 3).

Table 1. Comparison between different levels of wavelet decomposition

	Level 1	Level 2	Level 3	Level 4
Image 1: Brain	40.7	40.7	37.5	41.1
Image 2: Brainc1	43.18	43.18	41.1	43.71
Image 3: Brainc4	35.77	35.77	32.54	35.89
Image 4: Knee	36.34	36.34	36	38.38
Image 5: Shoulder	37.11	37.11	35.5	38.65

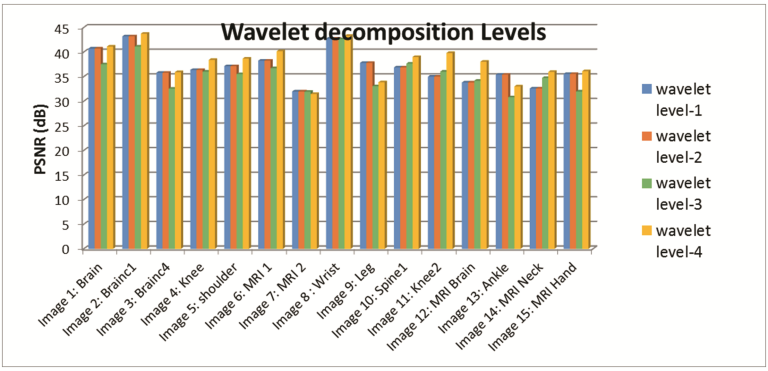


Fig. 3. Comparison of different levels of wavelet decomposition

The result shows that level 4 wavelet decomposition is giving better PSNR or visual quality for image due to exploiting more frequency divisions of an image (Table 2 and Fig. 4).

Table 2. Comparison between prediction techniques

	W/o prediction	Linear prediction	DARC prediction
Image 1: Brain	37.588	37.58	37.6
Image 2: Brainc1	40.79	40.78	41.1
Image 3: Brainc4	40.14	40.75	41.3
Image 4: Knee	36.63	37.34	41.5
Image 5: Shoulder	40.7	41.77	40.4

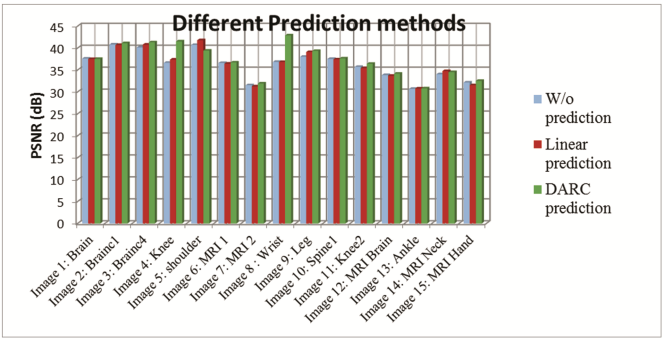


Fig. 4. Comparison between prediction techniques

The results show that DARC prediction technique is better. More clarity of reconstructed image can be achieved because DARC prediction considers more neighbouring pixels than the linear prediction.

The detail nature of High frequency band can be best exploited using Gaussian matrix which are pseudorandom values drawn from the standard normal distribution. The various Bernoulli matrix, Hadamard matrix, Toeplitz matrix and Gaussian matrix applied for LL band and results are given in Fig. 5 (Table 3).

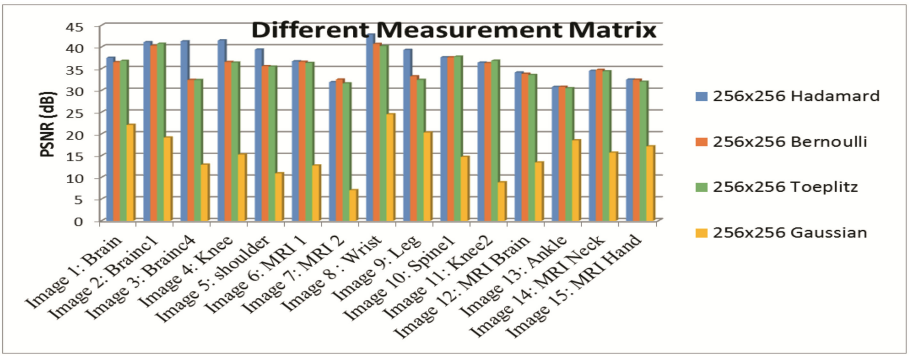


Fig. 5. Comparison of different measurement matrix

Table 3. Comparison between Bernoulli matrix, Hadamard matrix, Toeplitz matrix and Gaussian matrix applied for LL band

	Hadamard	Bernoulli	Toeplitiz	Gaussian
Image 1: Brain	37.5	36.5	36.8	22.06
Image 2: Brainc1	41.1	40.29	40.73	19.14
Image 3: Brainc4	41.3	32.41	32.37	12.9
Image 4: Knee	41.5	36.55	36.4	15.24
Image 5: Shoulder	39.4	35.57	35.46	10.9

Hadamard matrix is giving better results when used for measuring LL band. L1 minimisation works well with these measured values for recovering pixels (Table 4 and Fig. 6).

Table 4. Comparison of compression ratio without encoder and with encoder

	CR with LZW	CR with prediction	CR with CS
Image 1: Brain	1.57	1.21	1.17
Image 2: Brainc1	3.05	2.57	2.52
Image 3: Brainc4	4.33	3.54	3.5
Image 4: Knee	2.1	1.87	1.87
Image 5: Shoulder	5.45	5.27	5.00

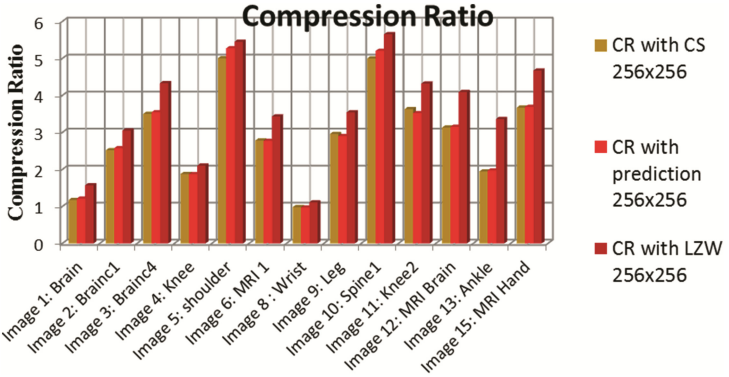
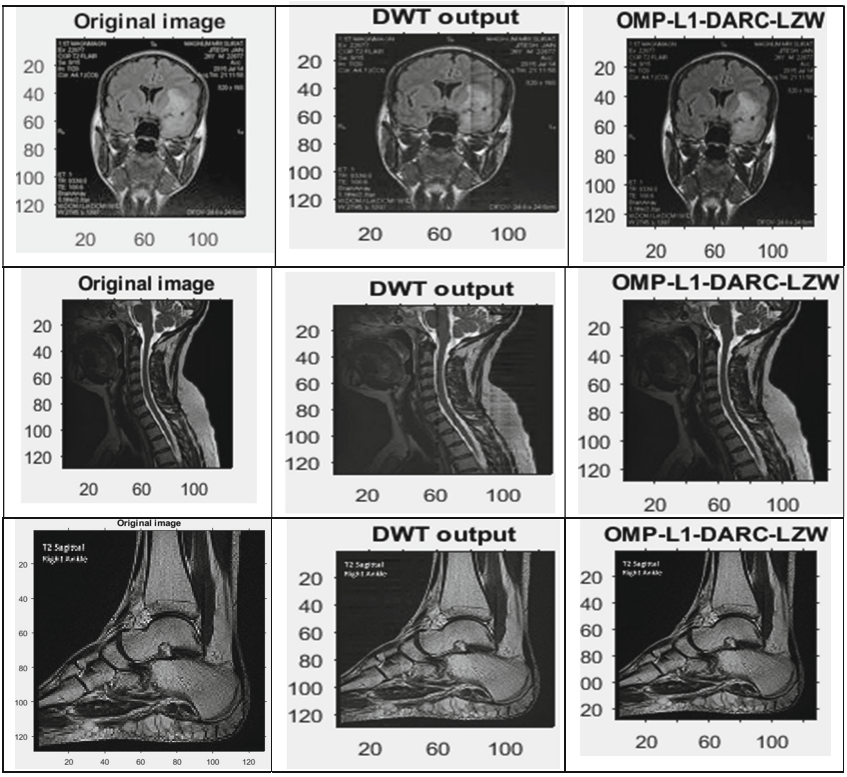


Fig. 6. Comparison of compression ratio on JPEG images

By this proposed method compression ratio is increased and thus less bandwidth only required for transmission through channel.

The visual quality is better when using the proposed method as shown in Table 5 when compared to CSPM technique used in [5] and all algorithms are tested on medical database collected from Hospital [8] and shown in result in above Table 5.

Table 5. Visual quality of images with DARC prediction and LZW encoder by using Hadamard matrix for measuring LL band and Gaussian matrix for measuring high bands.



5 Conclusion

The Compressive sensing measures small number of samples – in medical field this will help to reduce radiation time for MRI. Sparsity, incoherence and nonlinear reconstruction are three main components of CS.

Wavelet transform is proven sparsity domain for many signals. The high-frequency coefficients are sparse while the low frequency coefficients are not sparse. So, both bands processed separately and used different measurement matrix and recovery algorithms. The prediction technique helps in improving the PSNR.

For LL band, Hadamard matrix as measurement matrix, DARC as prediction method and L1 minimisation as recovery algorithm will result in higher PSNR. For high frequency bands, Gaussian matrix as measurement matrix, DARC as prediction method and OMP algorithm as recovery algorithm will result in higher PSNR and better reconstructed image.

The high PSNR, compression ratio and visual quality shows the proposed novel technique will be helpful in medical field especially for patients who undergoes MRI

scan. This proposed method will reduce the scan time but with better visual quality for easy diagnosing. Also the bandwidth reduction will be helpful in channel transmission.

Acknowledgments. **T**o all my friends and family who supported me to prepare this paper.

References

1. Qaisar, S., et al.: Compressive sensing: from theory to applications, a survey. IEEE J. Commun. Netw. **15**(5), 443–456 (2013)
2. Lustig, M., et al.: Compressed sensing MRI. In: IEEE Signal Processing Magazine, March 2008
3. Baraniuk, R.G.: Compressive sensing. In: IEEE Signal Processing Magazine, vol. 24(4), July 2007. Lecture notes
4. Ding, W., Yang, F., Pan, C., Dai, L., Song. J.: Compressive sensing based channel estimation for OFDM systems under long delay channels. IEEE Trans. Broadcast. **60**(2), 313–321 (2014)
5. Angayarkanni, V., Radha, S.: Design of bandwidth efficient compressed sensing based prediction measurement encoder for video transmission in wireless sensor networks. Wirel. Pers. Commun. **88**(3), 553–573 (2016)
6. Li, X., Bi, G.: Image reconstruction based on the improved compressive sensing algorithm. In: IEEE International Conference on Digital Signal Processing (DSP) (2015)
7. Candès, E.J., Wakin, M.B.: An introduction to compressive sensing. IEEE Sig. Process. Mag. **25**(2), 21–30 (2008)
8. Image database, Civil Hospital, Surat

Investigating Privacy Preserving Technique for Genome Data

Slesha S. Sanghvi(✉) and Sankita J. Patel

Department of Computer Engineering, Sardar Vallabhbhai National Institute of Technology, Surat 395007, Gujarat, India
slesha07@gmail.com, sankitapatel@gmail.com

Abstract. The rapidly growing genome sequencing technology has enabled the production of huge amount of sensitive genomic data. Presently a-days, it is conceivable to create highly detailed genotypes at lower cost. Sharing of genomic dataset is a key to comprehend the hereditary premise of human ailments. Because of the sharing of such information, genuine privacy challenges emerge with the expanded number of hereditary tests and immense gathering of such genomic information. The expanded accessibility of such information has real ramifications for individual protection, since it contains basic elements of human as well as contains, illnesses points of interest, insights about relatives, past and future era, responses to medication and substantially more.

To overcome the privacy issue in genomic data, previously some solutions had been purposed based on encryption techniques. However, the existing solutions has some limitations viz., identification of an individual from Genome Wide Association Study (GWAS) sets, generated test results contain Single Nucleotide Polymorphism (SNP) information about patients etc. In this work, we aim to propose a privacy preserving technique for genomic data that strengthen the security of genomic data.

Keywords: Genome sequencing · Privacy preserving · Disease detection
Privacy preservation for genome · Medical data security

1 Introduction

Genome fundamentally represents blueprint of a body. Our appearance, our maladies related data, our family history and much more information is determined by genome. Major applications of genome processing include recognizing criminals, prenatal testing and premature identification of diseases. Physically, it would require lots of time and efforts to establish the correlation between genomic information and human characteristics i.e. eye shading, obesity and so forth [1]. The healthcare can be revolutionized by medical data based on genome; however, the genomic data is also vulnerable against mishandle at the exact time. Essentially, genome data leads to social stigma, discrimination, employment and insurance denial. Human genome can put lifelong impact on an individual's life if it is leaked, as it is particularly stable. In literature, there exist numerous attempts that identify the risk of publishing genomic data [2, 3]. In Homer et al. [2], authors demonstrated that an individual could be

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 106–115, 2018.
https://doi.org/10.1007/978-3-319-73712-6_11

identified by using aggregate genomic data. Therefore, the data usage management of genomic data is crucial.

Genomes are basically acquired from a person in chemical form for digitization prepare. Both forms of genome i.e. digital and chemical; should be anticipated as it can be misused by some adversary. Due to inappropriate management, privacy breaching of genomic data can happen which leads to identity risk of individual. Therefore, genomic data must be handled carefully [4].

1.1 Motivation

Larger part of health care services requires genomic data to perform productive medical research or to perform certain diseases susceptibility test. This in turn requires sharing of genomic data to researcher or some third-party agency.

In addition, genetic data sharing among hospitals and research institutions is imperative for large-scale genetic studies. For example, let us consider that two medical organizations own the genetic datasets of their patients. The organizations need to run machine learning algorithms on the union of the datasets they own, without revealing their datasets to each other. Without utilizing a safe convention for these organizations to share data joint calculation on this information is infeasible. Other problem includes secure protocols for individual patient's disease susceptibility tests.

To address this issue, some previous work has been done, especially by Jha et al. [5]. Authors in [5] explore privacy preserving analysis for personal genomics. The idea is to utilize the outcomes found in Genome Wide Association Studies (GWAS) basically, to examine a particular disease susceptibility of an individual for getting a specific disease based on certain genetic markers that includes allele frequency and molecular markers. Limitation of this approach is mainly, it does not contain any secure way for examining an individual.

1.2 Contribution

As discussed, we focus on the problem of privacy preservation in Genome dataset. The problem is described as below:

> *Suppose, there is a data provider owing a private Genome dataset D. The dataset is required to be shared with various organizations for the two purposes viz. (1) medical research and (2) disease susceptibility test. The goal is to preserve the privacy of an individual, whose genome sequence is stored in D, while maintaining the accuracy of the results.*

To address the problem, we proposed a technique to preserve the privacy of an individual using Paillier cryptosystem and differential privacy. By using this technique an individual can generate his/her test results in secure way. Generated test results are then used by researchers to get statistical information without breaching individual's privacy.

In upcoming Sect. 2 we discuss about genomic background and issues related to securing genomic data. After that in next sections our proposed technique and its performance results are described.

2 Background

The genome of a human body contains set of genetic data that consist of mainly four different bases viz. Thymine (T), Guanine (G), Adenine (A), and Cytosine (C). A chromosome contains genetic data and these genes are accountable from form of functions dominant in human body all at once. There is distinction within the arrangement of those bases, which are supported by every DNA strand that results in individuation between individuals' genetic composition. Due to genetic differences, DNA of every person is different from reference genome by approximately 0.5%. SNP (Single nucleotide Polymorphisms) of a human body is commonest genetic dissimilarities [6].

2.1 DNA Sequencing and Analysis Process

DNA of the individual is collected from varied sample sources viz. skin, hair, saliva and blood. Once sample is collected, using extraction kit of DNA genetic data is extracted and then the process of sequencing that data is started using any sequencing platform. One of the widely used sequencing platforms is Illumina Sequencer [6]. For standard bioinformatics analysis, the digital DNA data is used after the sequencing process of DNA. In this manner, just physical securities are not sufficient to shield protection and supply wellbeing as computerized information is regularly replicated, changed and shared [6].

2.2 Features of Genome

After discussing about how genome information is handled, given us a chance to examine why genome data is delicate, and require more privacy than the standard medical data. Following are the major features of genomic data that creates privacy issues:

Genome contains sensitive information, which may bring about separation, work refusal and protection dissent, and mortification [4]. Immaterial intergenerational change i.e. DNA of individual changes less from one era to future era [7]. Likeness with blood-relatives i.e. one human genome contains loads of sensitive data regarding his blood-relatives. Closely connected peoples have very alike genomes [8]. Information contained by genome has various applications containing biomedical analysis, healthcare etc. [4]. By using partial genomic data, it is possible to get unavailable data i.e. it can leak disease information which is not even available. Human hereditary data contains six billion nucleotides which is very large in size.

2.3 Privacy Breach in Genome Data

Privacy breaching techniques of genomics are of three types:

Identity Tracing
In this kind of attack, intruder can build up a linkage between the data owner's hidden identity and an unidentified genome.

Attribute Disclosure Attack via DNA (ADAD)
Access of distinguished DNA by intruder and furthermore without utilizing express identifiers database that interfaces delicate properties with DNA-inferred information. These methods look at DNA information and give interface between the target's personality and its sensitive quality.

Completion technique
Intruder has just access of sanitized dataset without knowing about delicate area. In addition, intruder knows personality of genomic dataset however no get to. Point is to reveal the delicate area i.e. not a piece of genuine information [9].

2.4 Privacy Issues in Genome Data

Digital genomic data are used in various bioinformatics processes viz. searching on a genetic dataset, querying private data of genome, and sequence alignment. Regardless of their helpfulness in the medical field, these procedures present high danger of leakage of private data.

In first issue, insecure environment process of sequence alignment, DNA sequence alignment process demands high and expensive computation which is outsource to publicly available clouds. Yet, sending such kind of personal information to an open cloud may raise an issue as it is controlled by some third-party associations which create privacy issue [6].

In second issue, querying private data related to genome, as discussed, the human genome contains private data around an individual's science like whether an individual has a probability to build up a particular kind of disease. For the prevention of disease and furthermore for prescribing customized medicine, person's genome is taken and used to query against a list of known variations of disease to calculate susceptibility of diseases. In addition, to decide biological connections between persons, it is required to query genomic data to get result of tests like Paternity and ancestry [6].

In third issue, Private Data Sharing for GWAS, Genome-Wide Association Study (GWAS) defines connection between specific traits and common variations of genetics. From genomic records of thousands of individuals, GWAS examines Single Nucleotide Polymorphisms (SNP) and then it produces aggregate statistics. These aggregate statistics is then used to find connection between a disease and a SNP [10]. Homer et al. [2] presented an outline for robust and accurate detection of the existence of an individual by some known genotype in the mixture of complex DNA. The individual distance is measured from a test population and a reference. Then based on it t-test is calculated by using previously unknown individuals and the distance metrics to analyze this two populations and get the difference between them.

2.5 Related Work

As mentioned earlier for several types of privacy issues there are different techniques applied on genome data, but as shown in Table 1 all techniques having some of limitations. So, to improvise privacy for genomic data we proposed technique to resolve the above-mentioned issues.

Table 1. State of art in literature in privacy preserving techniques for Genome data

Privacy issues	Citation	Proposed solution	Limitations
Secure alignment on insecure environment	Chen et al. [11]	Use of seed-and-extend method	Higher data volume and high computational power
Querying private data related to genome	Alday et al. [12]	Use of Homomorphic encryption	Generated genetic tests results contains overall information regarding SNPs
Privacy preserving data sharing for GWAS	Fienberg et al. [10]	Uses ε-differentially mechanism	Releasing summary statistics for large data sets may not be enough to ensure the privacy of individuals

3 Proposed Technique

As discussed, we aim to preserve privacy for genome data. In the proposed technique, we utilize two significant methods, viz., Paillier cryptosystem [13] and differential privacy [14, 15]. The block diagram of our proposed approach is shown in Fig. 1.

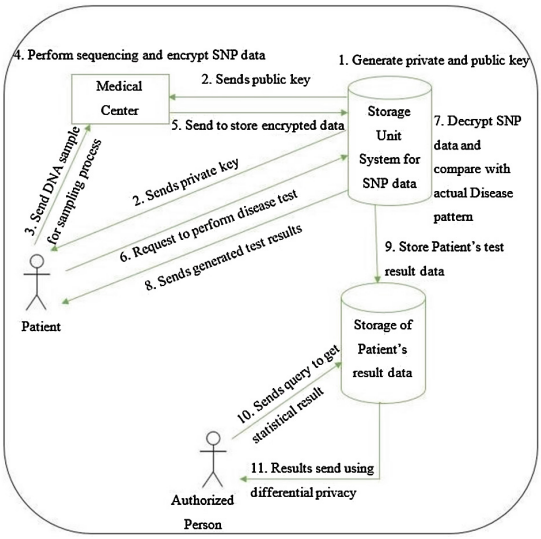


Fig. 1. Proposed technique

As shown in the Fig. 1, whenever patient enrolls for DNA sequencing process, patient gets private key. Using private key, patient will generate his/her test report by directly contacting the system. The generated results are then stored for statistical data analysis.

3.1 Generation of Secure SNP Data and Statistical Information

To achieve security, we apply Paillier cryptosystem on SNP data of the patient. Whenever patient enroll for DNA sequencing process at that time Storage Unit System (SUS) generates private and public key for the patient using Paillier cryptosystem. After the sequencing process, which is done at either medical center itself or by some third party, they will encrypt the SNP data of patient using patient's public key and store encrypted SNP data file of patient in SUS. If patient wants to go for any particular disease test, he/she will simply request to system. Patient forwards his/her private key to SUS. At SUS, system will perform the decryption operation. After decrypting the file, that file will be compared with particular disease's SNP file. For these tests, there are so many number of pattern files is stored at SUS. In this file, there is SNP information, which shows particular disease's pattern. By comparing this disease pattern file with patient's file test results are generated. After generation of test result, generated decrypted file of patient's information will be deleted automatically by the SUS. Moreover, generated test result will be forwarded to the patient.

Generated test results are stored in a file, which contains the basic information of patient i.e. patient id, sex, age and test result. This file will be further used for research purpose where researchers can send the queries to get some statistical information from database like, "How many number of patients having breast cancer who are male?" However, as explained before this generated result of query is prone to disclose the identity of patient. So, for statistical results we used differential privacy. Using Laplacian noise, we add noise into generated query result. After generating different statistical results, we add homomorphic encryption on these results by using Paillier cryptosystem.

4 Experimental Setup

In this section, we are going to discuss about experimental setup that we have created for Paillier cryptosystem and differential privacy. We have implemented Paillier cryptosystem in JAVA programming language and for differential privacy we have used R tool.

4.1 Paillier Cryptosystem Setup

Parameters to be used for key generation: We have taken public and private key of the length of 512 bits. To generate two random prime numbers p and q we have used 256 bits of length with certainty of 64 and used Random () function, this shows that randomly generated prime numbers are positive Big Integers that is probably prime with the length of 256 bits. Using p and q, we have generated n which is of 512 bits. For public key one more parameter is needed i.e. g. g is generated randomly using random function in the class of Z^{*2}_{n} of 512 bits. Based on p and q, lambda is generated. u is generated using g, n and lambda. We generated u direct at the time of decryption. So, patient having private key as lambda only.

4.2 Differential Privacy Setup

To preserve privacy in statistical results, we should add noise. So, for that ε value should be set as small as possible to get privacy of statistical results. On selected dataset, we perform number of cycle to get the value of ε at which we can get very minimum difference in between original dataset and by changing one of the row values of the dataset. We set threshold value as sum of all detected disease column. Number of time cycle we run, we set it as 1000. We took ε value in between 0 to 1 with the increment of 0.01. For adding noise, we have taken one parameter named as sigma who indicates $\Delta f/\varepsilon$ and its value is $1/\varepsilon$. As for our dataset sensitivity function carries value 1, which means Δf contains value as 1. We add Laplace noise over here as Lap $(1/\varepsilon)$. Also, we created one bound over here as added noise will be in a boundary of 0 to 2.

4.3 Dataset to Be Used

For genomic dataset, we take dataset from GWASdb. GWASdb is a database of SNP-phenotype association mapped to Disease Ontology and Human Phenotype Ontology [16]. These datasets basically contain the disease information related to chromosomes which affects DNA. These datasets stored at SUS, which will be used at the time of susceptibility testing process to compare it with patient's chromosomes.

After completing susceptibility test process generated dataset contains basic information related to patient i.e. patientID, sex, age, disease detected etc. For this we had taken dataset from UCI machine learning repository. Over here we used heart disease dataset of Hungarian data taken from this repository [17].

5 Performance Results

Paillier cryptography was performed on SNP data of patient to generate susceptibility test. The implementation of this cryptosystem was carried out on Eclipse IDE. For this we have taken heart disease chromosome details file from GWASdb which contains 11805 chromosomes. For encryption of patient's data, it takes approximately 201912 ms. And for susceptibility test checking it takes approximately 399764 ms in average case where no need to compare whole file for test results. In between we get the result as test is negative. But for worst case scenario where test results come as positive it takes approximately 990835 ms. Decryption operation is also performed at the time of susceptibility test process.

After susceptibility test, generated results are used for the purpose of research. On this dataset we fired below queries,

1. How many numbers of patients are having disease?
2. How many numbers of patients are having disease who are male?
3. How many numbers of patients are having disease who are female?
4. How many numbers of patients are having disease whose age group is between min and max (as per user enters)?
5. How many numbers of patients are having disease whose age group is between min and max (as per user enters) and gender (male or female)?

Generated ε values for above queries are shown in below Fig. 2, 3 and 4.

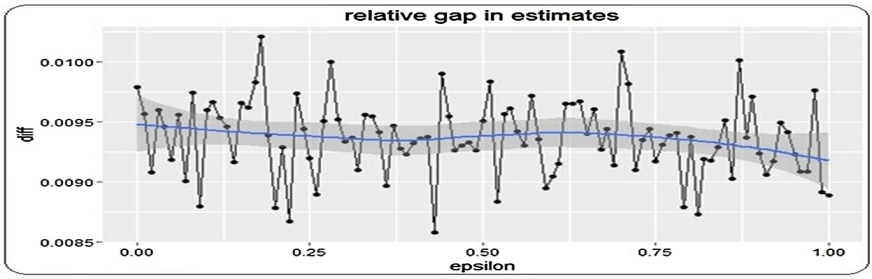


Fig. 2. Same graph for query 1, query 3, query 4 having age group 50–55 and query 5 with age group 50–55, gender female with ε value 0.43

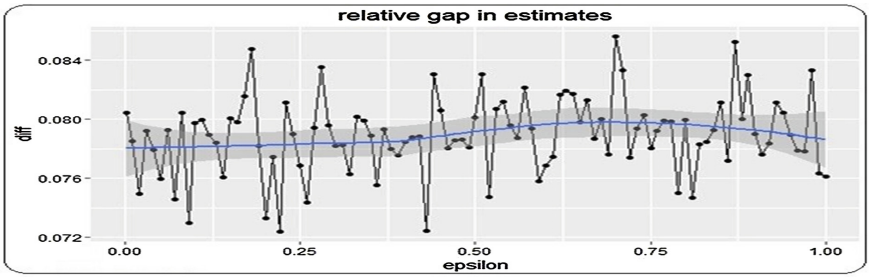


Fig. 3. Graph for query 2 with ε value 0.22

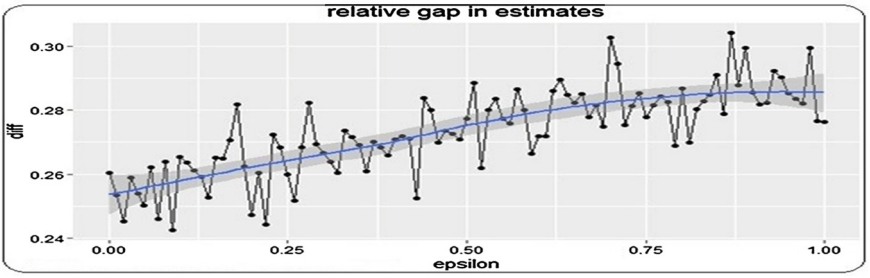


Fig. 4. Graph for query 5 having age group 50–55 and gender male with ε value 0.09

6 Result Analysis

6.1 Susceptibility Test Result Analysis

For susceptibility test of patients, we are having large amount of dataset so it takes more time for generating results. For comparison in between disease's chromosome file and patient's chromosome file its time complexity in best case is O(1) and in worst case scenario $O(n^2)$.

6.2 Statistical Result Analysis

As seen in the Figs. 2, 3 and 4, for epsilon values 0.43 we are getting lower difference. So, for this dataset of heart disease we can take ε as 0.43. In Fig. 3 we are getting two nearer values for ε i.e. 0.22 and 0.43, so we can take 0.43. But in Fig. 4 where our input is age in between 50 to 55 and gender having male we are getting privacy using very lower ε value that is 0.09. The actual result of this query is very low as, total number of patients having heart disease in between group of 50 to 55 and gender having male are only 3. So, from such kind of results, we can say that for different queries we are getting almost same values but for queries where we get very low population we got very less value of epsilon.

7 Conclusion and Future Work

As every year, bioinformatics field and sequencing process are becoming very important with increasing number of genomic tests. Persons use sequencing of DNA data for different number of aims and unwanted access to this sensible genetic information may create serious privacy breaching in the coming years. Due to the lack of techniques for privacy preservation it creates difficulties rather than benefits as there is revolutionary use of sequencing technology by medicine and health sciences.

With an aim to solve the privacy issue in genome data, in this paper, we mentioned the sensitivity of genome data and discussed various privacy breach techniques on genomic data. We proposed a technique based on the differential privacy and Paillier cryptosystem and discussed respective results.

In our work, after completion of susceptibility test there is no provision of encryption on generated result file. Future work could be to add encryption in result. The current work checks for a particular disease only that is, "Whether patient having X disease or not?" In future work, more feature can be added, using which, we can check and generate files containing details related to total number of chromosomes affected by particular disease. And then that statistical data would be store in a file with the use of differential privacy. And after performing susceptibility test patient can get the result as which diseases can be affected to them.

Major limitation of the proposed technique is the requirement of high computing power because there is a need for high number of chromosomes to be encrypted, decrypted and compared.

References

1. Genome-wide association studies. http://www.genome.gov/20019523. Accessed 10 June 2016
2. Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., Craig, D.W.: Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. **4**(8), 1000167 (2008)
3. Wang, R., Li, Y.F., Wang, X., Tang, H., Zhou, X.: Learning your identity and disease from research papers: information leaks in genome wide association study. In: CCS, pp. 534–544 (2009)
4. Naveed, M.: Hurdles for genomic data usage management. In: IEEE Workshop on Data Usage Management (DUMA), pp. 44–48, May 2014
5. Jha, S., Kruger, L., Shmatikov, V.: Towards practical privacy for genomic computation. In: Proceedings of the 2008 IEEE Symposium on Security and Privacy, pp. 216–230 (2008)
6. Akgün, M., Bayrak, A.O., Ozer, B., Sağıroğlu, M.Ş.: Privacy preserving processing of genomic data: a survey. J. Biomed. Inform. **56**, 103–111 (2015)
7. Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., et al.: Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science **328**(5978), 636–639 (2010)
8. Burdick, J.T., Chen, W.-M., Abecasis, G.R., Cheung, V.G.: In silico method for inferring genotypes in pedigrees. Nat. Genet. **38**(9), 1002–1004 (2006)
9. Erlich, Y., Narayanan, A.: Routes for breaching and protecting genetic privacy. Nat. Rev. Genet. **15**, 409–421 (2014)
10. Yu, F., Fienberg, S.E., Slavkovic, A.B., Uhler, C.: Scalable privacy-preserving data sharing methodology for genome-wide association studies. J. Biomed. Inform. **50**, 133–141 (2014)
11. Chen, Y., Peng, B., Wang, X., Tang, H.: Large-scale privacy-preserving mapping of human genomic sequences on hybrid clouds. In: NDSS (2012)
12. Ayday, E., Raisaro, J.L., Hubaux, J.-P.: Privacy-Enhancing Technologies for Medical Tests Using Genomic Data, Technical report (2012)
13. Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes. In: Stern, J. (ed.) EUROCRYPT 1999. LNCS, vol. 1592, pp. 223–238. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48910-X_16
14. Dwork, C.: Differential privacy. In: 33rd International Colloquium, ICALP 2006, Venice, Italy, Proceedings, Part II, 10–14 July 2006
15. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: privacy via distributed noise generation. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 486–503. Springer, Heidelberg (2006). https://doi.org/10.1007/11761679_29
16. GWASdb SNP-Disease Associations dataset. http://amp.pharm.mssm.edu/Harmonizome/dataset/GWASdb+SNP-Disease+Associations. Accessed 10 June 2016
17. UCI machine learning database. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.hungarian.data. Accessed 10 June 2016

Dimensionality Reduction Using PCA and SVD in Big Data: A Comparative Case Study

Sudeep Tanwar[1(✉)], Tilak Ramani[1], and Sudhanshu Tyagi[2]

[1] Department of CE, Institute of Technology, Nirma University, Ahmedabad, India
{sudeep.tanwar,16mcei19}@nirmauni.ac.in
[2] Department of ECE, Thapar University, Patiala, Punjab, India
s.tyagi@thapar.edu.in

Abstract. With the advancement in technology, data produced from different sources such as Internet, health care, financial companies, social media, etc. are increases continuously at a rapid rate. Potential growth of this data in terms of volume, variety and velocity coined a new emerging area of research, Big Data (BD). Continuous storage, processing, monitoring (if required), real time analysis are few current challenges of BD. However, these challenges becomes more critical when data can be uncertain, inconsistent and redundant. Hence, to reduce the overall processing time dimensionality reduction (DR) is one of the efficient techniques. Therefore, keeping in view of the above, in this paper, we have used principle component analysis (PCA) and singular value decomposition (SVD) techniques to perform DR over BD. We have compared the performance of both techniques in terms of accuracy and mean square error (MSR). Comparative results shows that for numerical reasons SVD is preferred PCA. Whereas, using PCA to train the data in dimension reduction for an image gives good classification output.

Keywords: Dimensionality reduction · Principle component analysis
Singular value decomposition · Big data

1 Introduction

Volume of data is increasing exponentially to Tera byte or Peta byte from many sources like biomedicine, social media, Internet of Things (IoT), etc. All data on the planet is growing 40% a year. International data corporation (IDC) has predicted that volume of data will grow above 40 ZB by 2020 [1]. The comparative growth of digital data over time (measured in years) is shown in Fig. 1(a) which is indicates, In 2013 digital universe had 5500 EB, but in 2020 it will be 44 ZB, a 10-fold increment in very short span of time. The top three sources of data are sales & financial transactions (56%), leads & sales contacts from customer databases (51%), and email & productivity applications (tied at 39%). Almost a quarter of respondents (19%) are managing less than a tera byte of data, while only 7% are managing more than a peta byte. Although the average company

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 116–125, 2018.
https://doi.org/10.1007/978-3-319-73712-6_12

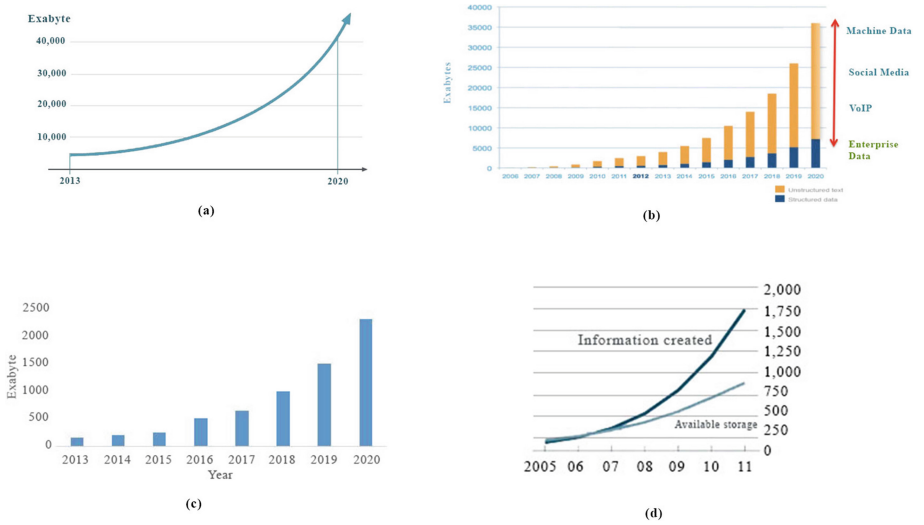


Fig. 1. (a) World wide growth of digital data, (b) Data growth in Enterprise, (c) Data & storage increase over the years, (d) Growth of data in heath care sector

manages 162.9 TB of data, the average enterprise has 347.56 TB of data [1], which is increasing by 33% a year. Figure 1(b) shows the incremental growth in structured and unstructured data over the years. Health care data covers large segment of entire digital universe, and it is increasing 48% in a year. All data in the health care was 153 EB in 2013, but it is expected to be 2,314 EB in 2020. As data volume is growing exponentially, available storage to accommodate it also need to be updated accordingly. Comparative study of growth in data and corresponding storage is shown in Fig. 1(d), indicating that storage is not increasing as rapidly as data. This exponentially increment in data is very complex in several situations like to maintain (a) the real time monitoring for (i) health sector (ii) car parking system (iii) fire alarms, (b) security of (i) offices, (ii) hospitals (iii) defense area and many more. Currently available computing infrastructure and analytical algorithms are not able to manage and process the current form of generated BD. In some situations this data is redundant too, therefore, cleaning of data is required to maintain high quality. Compared to raw data, this cleaned data is very small in size but has important information. To clean this raw data, we have used DR techniques in this paper.

DR is the procedure to convert a dataset have vast number of dimensions into a data subset with less dimensions ensuring no lose of important information. The importance of DR is to improve the accuracy of prediction of classifier, and to decrease the cost of computation. These techniques are basically used to solve machine learning problems to get quality features in classification and regression. Some advantages of DR are summarized as under:

- It compresses data and reduces the storage requirements.
- It reduce the computation time.

– It considers multi-collinearity that gives better performance of the model.
– It eliminates redundant features.
– It helps in eliminating the noise.

To understand the concept of DR we have selected PCA and SVD two different techniques. These techniques are investigated thoroughly, and compared by executing with machine learning algorithm. PCA takes a dataset comprising of the set of tuples focusing on points lying on a high-dimensional space. PCA also searches for the directions with which the tuples line up best. Main objectives of PCA are:

– Form a data table it extricate the vital information.
– Keeping the vital information only, it compress the size of the dataset.
– To simplify description of the dataset.
– To analyze the structure, and factors.

The idea behind PCA is to consider a matrix M be the set of tuples and search for the eigen vectors of MM^T or M^TM. The axis related to the first eigen vector, the one along with which the variance of raw data is maximized. Now, one can apply this transformation to that data. Similarly, the axis related to the second eigen vector is the axis along with which the variance of distances from the first axis is most prominent, and so on. Hence, one can say that PCA is a data mining process. The high dimensional data is supplanted by the projection on essential axes. These axes are related to the largest eigenvalues. Finally, raw data is estimated by data that has less dimensions compared to raw data.

On the other side, SVD is a method to distinguish the dimensions along with which data points show the highest variation. SVD permits to get the best estimation of the raw data using less dimensions. This approach permits a correct portrayal of any matrix. Furthermore, this approach removes the less essential dimensions of that portrayal to create an approximate portrayal with any coveted dimensions. SVD decompose an $m \times n$ matrix, M into U, S, and V. This decomposition has the form USV^*. Here, U is an $m \times r$ matrix, S is a $r \times r$ diagonal matrix, & V is an $n \times r$ matrix. We can utilize them to diminish the number of vectors to the variance we actually required. Diminishing the number of vectors can remove noise from the raw dataset.

1.1 Research Contribution of Paper

Contributions of this paper are as follows:

– We have reduced the dimension of sparse and dense dataset using PCA and SVD.
– We have compared the performance of PCA and SVD by applying them on two different dataset.

The rest of the paper is structured as follows. Section 2 highlights previous work done by researchers in this domain with pros and cons of individual. Section 3 highlights the need of DR and present the techniques PCA and SVD. Section 4 presents the comparison result of both techniques in terms of accuracy & mean square error and finally Sect. 5 concluded the paper.

Table 1. Comparison of existing approaches

Author	Problem statement	Solution	Drawback
Swati *et al.* [2]	The classification of high dimensional data give wrong outcomes	A method that utilizes DR techniques	Another classifier for classification can be used instead of ARTMAP to reduce more time
Person *et al.* [4]	Show points in plane or higher dimensional space by the straight line or plane	Principle component analysis	It becomes more cumbersome when we have more variables which involves the determination of least root
Oja *et al.* [7]	PCA for neural networks	A completely parallel (nonhierarchical) design that gets orthogonal vectors spanning an m-dimensional PCA subspace	Lateral connections between the units are not considered
Sanger *et al.* [8]	Measure the data in network results can be troublesome without exact learning of the distribution on the input data	Optimality principle for training an unsupervised feedforward neural network	The algorithm is only for single-layer linear networks
Henry *et al.* [13]	Identify the dimensions along which data points	Singular value decomposition	When there is no change in one of the axes, SVD fails
Deerwester *et al.* [14]	Dimensionality reduction issue with regards to information retrieval	Use SVD for making features representing multiple words and after that comparing them	Implementation issues will emerge as in raw vector methods, the estimation of such retrieval improving methods must be reevaluated
Sarwar *et al.* [17], Brand *et al.* [16]	The high cost of finding the SVD	Update an existing SVD without recomputing it from scratch	Works well for some recommender applications and less well for others

2 Related Work

This section highlights the work done by various researchers in this domain. Swati *et al.* [2] classified the high dimensional raw data that creates incorrect outcomes. To obtain precise outcomes, high dimensional raw dataset should be compressed to enhance the accuracy of outcome. Repetitive and the conflicting data should be eliminated to achieve it. In [2] authors have presented a constraint selection algorithm to utilizing DR techniques. Because of the DR techniques the computation time is reduced. Tarun *et al.* [3] took the DR technique diminish space and improves the overall performance. For DR meta-heuristics techniques were utilized. To reduce the space DR technique is more valuable; fast information retrieval, optimized image processing, good visualization, exact classification for area oriented datasets. PCA for DR was introduced by Pearson *et al.* [4] and modern

representation was given by Hotelling *et al.* [5]. Selection of the dimensions using PCA was explained by Jolliffe *et al.* [6]. One dimensional PCA was implemented for neural networks by Hebb learning *et al.* [7] and later on extended to hierarchical multidimensional PCA by Sanger [8–10]. Further, in [7] authors have given a completely parallel plan that concentrates on orthogonal vectors traversing an m-dimensional PCA subspace. Baldi *et al.* [11] demonstrated the error surface for linear, three layer auto-associators with hidden layers of width m has global minima relating to input weights that traverse the m-dimensional PCA subspace.

SVD was first introduced by Golub *et al.* [12] and later on Henry & Hofrichter [13] utilized it to recognized the dimensions along which data points shows the largest variation. Deerwester *et al.* [14] analyzed the DR issues with regards to information retrieval. They were compared documents using the words they consist of, and they proposed a method of producing features representing different words and then comparing them. Recently, Sarwar *et al.* [15] used SVD for recommender systems. One of the difficulties of utilizing an SVD-based algorithm for recommender systems is the high cost to search the SVD. In spite of the fact that it can be computed off-line, finding the SVD can in any case be computationally intractable for vast databases. To solve this issue, various researchers have analyzed incremental techniques that changed current SVD without recomputing it from scratch [16,17]. Table 1 show the details of several proposals.

3 Dimensionality Reduction

We live in the age of BD where we do not have just a handful observations and variables; possibly often hundreds or even thousands of variables that we need to analyze, identify important trends, patterns, and to gain some insights about the businesses or for profit organizations to make policy decisions or even to do some basic research. Hence, we have many variables against which we have many observation stored in the same table. Now problem is how out of many observations select smaller group that contains chunk of observations. On the other hand we might be overwhelmed by the sheer number of variables in the data sets and some variables further more may be highly correlated or highly similar to each other creating additional problems with their interpretation and modeling itself. Hence, we might be interested to reduce the number of variables.

Second issue, we might be interested to revolves the way too many variables within our data sets and we're interested to see how our variable hang together, and how they can describe the datasets in the most efficient way. The variables may described very similar things and we're looking for the underlying similarity. Then group those variables together into a single broad dimensions that will describe our data set most efficiently. It is not advisable to enter all the variables in a single model because it's very often quite inefficient, computationally expensive, and their are high correlations among variables. PCA is especially helpful in this situation.

To reduce the dimensions of data apply cluster analysis over it. Further to reduce the dimensions of constructs, PCA and exploratory factor analysis give

good results. In this paper, we have discussed the reduction of dimension of constructs or reduction in number of variables in existing data set. Next subsection present the PCA in detail.

3.1 Dimensionality Reduction Through Principle Component Analysis

PCA is a technique for extracting important factors (components) from a vast set of variables accessible in a dataset. It extricates low dimensional set of elements from a high dimensional dataset with an objective of getting as much information as possible. With a less factors, representation it turns out to be significantly more important. PCA is more valuable when managing three or more dimensional data. It is always performed on a symmetric correlation or covariance matrix. This implies that the matrix out to be numeric and have standardized data. First principal component is a linear combination of original predictor factors which catches the highest variance in the dataset. It decides the direction of most variability in the data. Higher the variability caught in first component implies more information caught by component. No other component can have variability higher than first principal component. The first principal component brings out to be a line which is nearest to the data i.e. it limits the sum of squared distance between a data point and the line. Likewise, we can also compute the second principal component. Second principal component is a linear combination of original predictors like first component which catches the rest of variance in the dataset and is uncorrelated with the first principal component outcome. That is, the correlation between first and second component should be zero. The direction of two components are orthogonal, if they are uncorrelated.

All succeeding principal component follows a similar idea, they catch the rest of variations without being correlated with the past component. The directions of these components are distinguished in an unsupervised way that means, the response variable is not used to decide the component direction. Thus, it is an unsupervised approach. As an example, M is a matrix, rows of which refers to the point in space, we can compute $M^T M$ and eigen pairs of that point. E, the matrix, which columns as the eigen vectors, ordered in such a way that largest eigenvalue comes first. Let the matrix L having the eigenvalues of $M^T M$ along the diagonal, in such a way that largest value comes first and 0's in other entries. Then, though $M^T Me = \lambda e = e\lambda$ for every eigen vector e and its related eigen value λ, it is understandable that:

$$M^T ME = EL \tag{1}$$

It has been observed that ME is the points of M changed into another coordinate space, in which, the first axis that is related to the largest eigen value, is critical. The variance of points along that axis is the most. The second axis, related to the second eigen pair, is the next noteworthy in the similar way, and this pattern proceeds for every eigen pairs. If it is desired that, M is transformed into a space having less dimensions, then the choice having the most important uses the eigen

vectors related to the highest eigen values and discards the other eigen values, i.e., if E_k is the first k columns of E, then ME_k is the k-dimensional potrayal of M. Next subsection presents another DR technique, that is Singular Value Decomposition.

3.2 Dimensionality Reduction Through Singular Value Decomposition

SVD permits a accurate portrayal of any matrix, and furthermore SVD makes it simple to remove the less vital factors of that portrayal to deliver an approximate portrayal with any coveted number of dimensions. M is an $m \times n$ matrix The rank of M is r. Where the matrix rank r is the largest number of rows or columns that we can get for nonzero nonlinear combination of the rows which is the all-zero vector 0, in other words, a set of these rows or columns is independent of each other. Then,

- U be $m \times r$ column-orthonormal matrix. Each columns of this matrix is a unit vector and the dot product of any two columns is 0.
- V be $n \times r$ column-orthonormal matrix. V is utilized as its transposed form, so that the rows of V^T that are orthonormal.
- S be a diagonal matrix. Elements, that are not on the main diagonal are 0. S elements are known as the singular values of M.

If we take a very large matrix M by SVD components U, S, and V, however these three matrices are also extensive to store. Then,

$$M_{m\times n} = U_{m\times r} S_{r\times r} (V_{n\times r})^T \tag{2}$$

To diminish the dimensionality of the three matrices, the most ideal approach can be set the singular values that are smallest to zero. We can remove s columns of U and V, if the s smallest singular values are set to 0. Advantages of using SVD are:

- SVD gives best axis to project on, means, minimum sum of projection error.
- Minimum construction error.

But at the same time SVD have some gaps, which are:

- **Interpretability problem:** A singular vectors specifies a linear combination of all input columns and rows.
- **Lack of sparsity:** Singular vectors are dense.

4 Result and Discussion

In this section, we have compared PCA and SVD in terms of accuracy and mean square error. PCA works by finding the eigenvectors of the covariance matrix and ranking them by their respective eigenvalues. The eigenvectors with

the greatest eigenvalues are the principal components of the data matrix. The matrix of eigenvectors in PCA are the same as the singular vectors from SVD, and the eigenvalues generated in PCA are just the squares of the singular values from SVD. While formally both solutions can be used to calculate the same principal components and their corresponding eigen/singular values, the extra step of calculating the covariance matrix in PCA can lead to numerical rounding errors when calculating the eigenvalues/vectors. Moreover, PCA gives the subspace that spans the deviations from the mean data sample as output, and SVD provides a subspace that spans the data samples themselves (or, a subspace that spans the deviations from zero).

4.1 Comparison of PCA and SVD in Terms of Accuracy

We have considered multivariate "Spam E-mail Dataset", of UCI Machine Learning Repository [18]. Before applying DR techniques accuracy was 93%. Here, our ultimate objective is to compare performance of PCA and SVD. Figure 2(a) show as number of attribute decreases, accuracy of PCA and SVD decreases. For some number of attributes, PCA gives maximum accuracy, but then drops drastically. But in SVD accuracy decreases gradually with decrease in attributes. It is important to note that PCA (5–7 min) takes lot of time compared to SVD (in seconds) to process around four thousands records. We have considered another dataset, "Wisconsin Breast Cancer Dataset" from UCI repository [18]. We have performed Same steps as performed in previous dataset, to analyse the performance of PCA and SVD. Initially the accuracy of data set was 97.54%. For this dataset, as number of attributes decreases, first accuracy of SVD increases, but then decreases gradually. But for PCA, accuracy decreases dramatically as number of attributes decreases, as shown in Fig. 2(b).

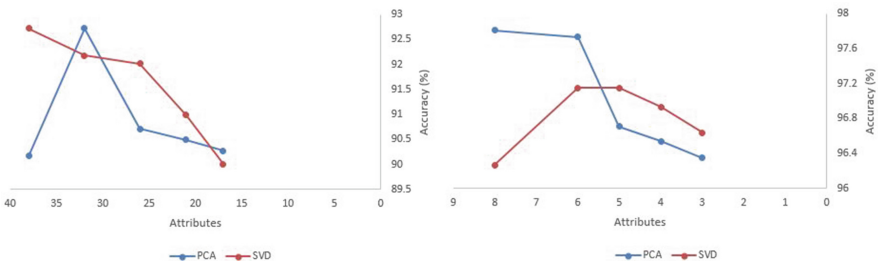


Fig. 2. (a) Spam E-mail dataset accuracy (%) Vs. no. of attributes, (b) Wisconsin breast cancer dataset accuracy (%) Vs. no. of attributes

4.2 Comparison of PCA and SVD in Terms of Mean Square Error

We have also compared PCA and SVD in terms of mean square error. For "Spam E-mail Dataset", mean square errors for PCA and SVD are almost same as shown in Fig. 3(a). For "Wisconsin Breast Cancer Databset" mean square errors for SVD is more than PCAb, as shown in Fig. 3(b).

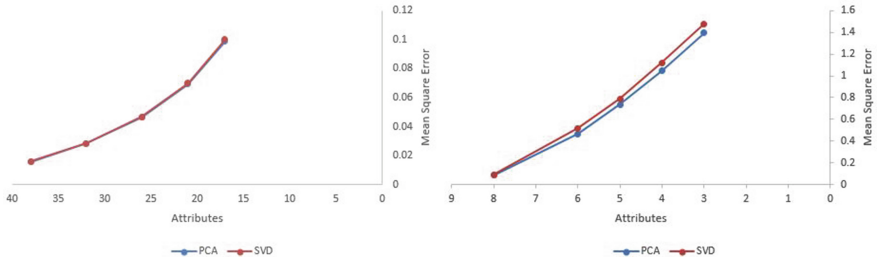


Fig. 3. (a) Spam E-mail dataset mean square error Vs. no. of attributes, (b) Wisconsin breast cancer dataset mean square error (%) Vs. no. of attributes

5 Conclusions

Their is an urgent requirement to process rapidly generated data with less storage space. Moreover this data is uncertain, redundant and inconsistent. Therefore, DR techniques comes in to picture, for fast processing of this data. Their are many approaches exist in the literature for DR, but we have discussed two of them, Principle Component Analysis and Singular Value Decomposition. We have compared the performance of both in terms of accuracy and mean square root. From comparison we have concluded that through SVD we get the "effective dimensionality" of a set of points. Moreover, for numerical reasons, it is preferred to use SVD. As it doesn't need to compute the covariance matrix which can introduce some numerical problems. Because there are some pathological cases where the covariance matrix is very hard to compute. So the SVD is numerically more efficient. Using the SVD to training data to diminish the dimension in an image gives good classification output. In future we will explore more DR approaches and apply tensor decomposition over these.

References

1. Gantz, J., Reinsel, D.: IDC, The Digital Universe (2014)
2. Swati, A., Ade, R.: Dimensionality reduction: an effective technique for feature selection. Int. J. Comput. Appl. **117**(3), 18–23 (2015)
3. Gupta, T.K., et al.: Dimensionality reduction techniques and its applications. J. Comput. Sci. Syst. Biol. **8**(3), 170 (2015)
4. Person, K.: On lines and planes of closest fit to system of points in space. Philos. Mag. **2**, 559–572 (1901)
5. Hotelling, H.: Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. **24**(6), 417 (1933)
6. Jollie, I.T.: Principal Component Analysis. Springer, New York (1986)
7. Oja, E.: Simplifed neuron model as a principal component analyzer. J. Math. Biol. **15**(3), 267273 (1982)
8. Terence, D.: An optimality principle for unsupervised learning. In: NIPS, pp. 11–19 (1988)

9. Kung, S.Y., Diamantaras, K.I.: A neural network learning algorithm for adaptive principal component extraction (APEX). In: International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1990, pp. 861–864 (1990)
10. Rubner, J., Tavan, P.: A self-organizing network for principal-component analysis. EPL (Europhysics Letters) **10**(7), 693–696 (1989)
11. Baldi, P., Hornik, K.: Neural networks and principal component analysis: learning from examples without local minima. Neural Netw. **2**(1), 53–58 (1989)
12. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. JHU Press, Baltimore and London (2012)
13. Henry, E.R., Hofrichter, J.: Singular value decomposition: application to analysis of experimental data. Methods Enzymol. **210**, 129–192 (1992)
14. Deerwester, S., Harshman, R., et al.: Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. **41**(6), 391–397 (1990)
15. Sarwar, B., et al.: Application of dimensionality reduction in recommender system-a case study. Technical report, DTIC Document (2000)
16. Brand, M.: Fast online SVD revisions for lightweight recommender systems. In: Proceedings of the International Conference on Data Mining, pp. 37–46. SIAM (2003)
17. Sarwar, B., et al.: Incremental singular value decomposition algorithms for highly scalable recommender systems. In: Fifth International Conference on Computer and Information Science, pp. 27–28 (2002)
18. Lichman, M.: UCI Machine Learning Repository (2013)

A Comparative Analysis of Ionospheric Effects on Indian Regional Navigation Satellite System (IRNSS) Signals at Low Latitude Region, Surat, India Using GDF and Nakagami-m Distribution

Sonal Parmar[1,2(✉)], Upena Dalal[2], and Kamlesh Pathak[3]

[1] Department of Electronics and Telecommunication, M.P.S.T.M.E., NMIMS University, Mumbai, India
sonal.parmar@nmims.edu
[2] Department of Electronics Engineering, SVNIT, Surat, India
[3] Applied Physics Department, SVNIT, Surat, India

Abstract. Indian regional navigation satellite system (IRNSS) is our own navigation system designed to provide various navigational and timing services in Indian region. The present paper discusses ionospheric scintillation effects on IRNSS signals in Surat, (21.16°N, 72.68°E; Geomagnetic 12.90°N, 147.35°E) India, a low latitude station. Ionospheric scintillations are one of the major sources of errors in satellite communication which may results in loss of lock of with particular satellite causing discontinuity of satellite services. Ionospheric Scintillation is experienced by satellite signals when propagating through various layers of atmosphere in terms of random fluctuations in amplitude and phase of signals, and also causes ionospheric delay. The present analysis is done on 3rd October 2015 data for calculation of Ionospheric scintillation measuring parameters like total electron content (TEC) and Vertical TEC using IRNSS data. During this time only four satellites were launched from PRN 1 to 4 by Indian space research organization (ISRO). The comparative behavioral analysis of TEC variation is done using Gaussian distribution function (GDF) and Nakagami m (NGK) model distribution. The carrier to noise (C/N) ratio and elevation angle variations for satellite PRN numbers from 1 through 4 is also carried out. It is seen from comparative analysis that total electron content variations in this geographical location are more following the Nakagami distribution as compared with GDF. The results can be utilized further for developing model for analyzing variations in TEC in this region.

Keywords: Errors in satellite communication · GDF · GNSS
Ionospheric scintillation · Ionospheric scintillation parameters · IRNSS signals
Nakagami distribution · Satellite navigation

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 126–136, 2018.
https://doi.org/10.1007/978-3-319-73712-6_13

1 Introduction

The Indian Regional Navigation Satellite System (IRNSS) is an independent navigation system which is developed by Indian Space Research Organization (ISRO), India. This navigation system is going to provide many applications related to navigational (standard positioning service and restricted service) and timing services in India as well as fifteen hundred kilometers around the Indian region. It is comprising of three segments like Space segment, ground segment and User segment. The space segment is comprising of seven satellites starting from IRNSS 1A to 1F, out of which three satellites are launched in geostationary earth orbit (GEO) at 32.5°, 83° and 131.5° East and four in Geosynchronous orbit (GSO) at inclination of 29° with longitude crossing at 55° and 111.75° East. The ground segment consists of various master and controlling earth stations keeping track of various orbital parameters of satellite location in orbit. User segment is comprising of IRNSS receiver which will be able to receive satellite signals decode them and able to provide useful navigational and timing information to user.

One of the IRNSS receiver granted by ISRO, Ahmedabad, is located at SVNIT, Surat which is situated at low latitude and equatorial anomaly region of India is as shown in Fig. 1. Hence it is more interesting to investigate the effects caused on IRNSS signals when it is propagating through various layers of atmosphere in this region. Figure 2 is illustrating sky plot of seven satellites starting from Pseudo Random number (PRN) 1 through 7 over the SVNIT region.

Satellite signals when propagating through various layers of atmosphere experience various sources of errors. Ionospheric scintillation is one of the major source of error for satellite signals. It is phenomena in which there can be sudden fluctuation in the satellite signals, there can be phase variations and amplitude scintillations. This can lead to loss of lock and cycle slip with respective satellite because of degradation in carrier to noise (C/N) ratio of the signals and hence can disrupt all services offered by the satellite during that period. This has led to scintillation studies in which various

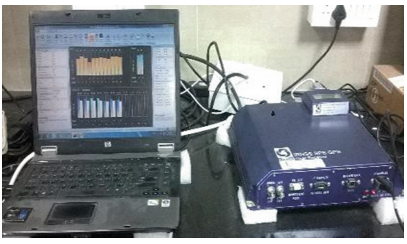


Fig. 1. Set up of IRNSS receiver at SVNIT, Surat.

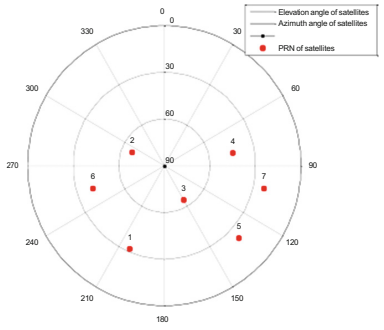


Fig. 2. Sky plot of IRNSS Constellation over SVNIT, Surat

researchers are analyzing the ionospheric scintillation effects on satellite signals and how it can be minimized.

This paper also discusses ionospheric scintillation effects on IRNSS signals. Ionospheric scintillation effects are very fast and random fluctuations of satellite signals in terms of variations in amplitude and phase caused by small scale irregularities present in the Ionosphere. This effects varies form geographical location, local time, season, solar radiations and magnetic activity as mentioned by Kintner et al. [1].

Paper is organized as follows. Section 2 is reviewing related work with IRNSS & Ionospheric Scintillation. Data and its analysis is presented in Sect. 3. Results are presented in Sect. 4. Finally discussions are concluded in Sect. 5.

2 Related Work

IRNSS receiver can be single frequency or dual frequency receivers operating in two frequencies namely L5 (1176.45 MHz) and/or S (2492.028 MHz) band. Modulation schemes used for signals are Binary phase shift keying (BPSK-1) and Binary offset carrier (BOC 5, 2) as mentioned by Saini and Gupta [2]. IRNSS receiver designed by ISRO is also capable of receiving Global Positioning System (GPS) signals. The advantages of our Indian navigational satellite programs developed by India like Geo augmentation system (GAGAN), IRNSS and INSAT-MSS system for the civilian use in Srilanka region is also discussed by Senanayake [3].

The IRNSS service is also beneficial to Australian continent as mentioned by Zaminpardaz et al. [4]. This paper provides insights about stand-alone positioning services over Australia. The SPS signals Pseudo random codes (PRN) are accomplished by Yashaswini [5]. It also highlights use of PRN gold codes by using Matlab, Xilinx, ISE simulator and Sparton FPGA environments. The Geometric dilusion of Precision (GDop) is very important parameter for identifying satellite geometry and finding range error in meters as described by Sekhar et al. [6].

A software designed receiver FGI-GSRx was developed by Thombre et al. for tracking IRNSS signals in northern Europe as mentioned in [7]. The receiver was able to carrier to noise power ratio for three IRNSS and GPS satellites. It has been shown by Majithiya et al. that IRNSS signal has group delay differential correction parameter to correct for L5 and S band as presented in [8]. Kalman filter based approach can be used to track the loop of Phase locked loop and provide better performance of receiver under ionospheric scintillation effects as discussed by Manjula et al. [9]. The analysis of position accuracies provided by various navigation system like GPS, IRNSS and hybrid are done by Manjunatha and Kiran [10]. Kumar et al. [11] has discussed the estimation of satellite geometry of IRNSS in terms of absolute, relative velocity, psuedorange and carrier phase.

Ionospheric scintillation effects on satellite or GPS signals results into Signal refraction and Signal diffraction [12]. Ionospheric scintillation measuring parameters are Amplitude scintillation index known as S4, Phase scintillation index $\sigma\phi$, Percentage of Scintillation occurrence (PSO), Total electron content (TEC) or Slant TEC, Vertical TEC (VTEC), Rate of Change of TEC (ROT), Satellite Elevation angle, Lock time, as mentioned by various researchers in [13–15]. Ionospheric scintillation studies

for GPS signals for a particular day for diurnal study is done by Parmar et al. [16]. Ionospheric scintillation effects on IRNSS signals using GDF and Nakagami distribution is also investigated by Parmar et al. [17]. Thus in this section related work done by various researchers in area of IRNSS and Ionospheric scintillation effects on GPS and satellite signals are discussed.

3 Data and Analysis of Work

The data and work done here is collected and received by IRNSS UR receiver which is installed at SVNIT, Surat. The receiver is capable of tracking all PRN satellites at L5 and S band. Measurement and calculation of different ionospheric scintillation parameters is done in MATLAB using the 3 Oct 2015 data and by using mathematical formulation as presented by Tanna et al. for the same region in [14]. There are various ionospheric scintillation parameters but out of them total electron content is considered here for analysis purpose. Earlier studies have been done for GPS signals.

3.1 Calculation of Total Electron Content (TEC) or Slant TEC (STEC)

The speed of propagation of satellite signals depends upon number of free electrons in its path, known as total electron content (Slant TEC), the number of free electrons in a tube of 1 m^2 cross section extending from the receiver to the satellite as mentioned by Misra and Enge [12]. One TECU is 10^{16} electrons per m^2.

$$TEC = \int_{S}^{R} n_e\ (l)dl \tag{1}$$

Where n_e (l) is the variable electron density along the signal path, and the integration is along the signal path from the satellite to the receiver. TEC varies typically between 1 and 150 TECU. TEC is calculated here using the Eq. (2), where for IRNSS receiver, **f1 = (L5 band frequency = 2492.028 MHz) and f2 = (S1 band frequency = 1176.45 MHz)**, P1 and P2 are Pseudo ranges of each frequencies f1 and f2 respectively.

$$TEC = \left[\frac{f_1^2 * f_2^2}{f_2^2 - f_1^2}\right] \frac{(P_1 - P_2)}{40.3} \tag{2}$$

3.2 Calculation of Vertical Total Electron Content (VTEC)

The STEC depends upon the signal path geometry when it is propagating through various layers of ionosphere it is important to calculate VTEC which is not depending to upon the elevation of signal path. TEC variations forms the electron irregular patches in the ionosphere. And when the satellite signal propagates through these patches;

Ionospheric scintillation results. The VTEC is calculated using the Eq. (3) as mentioned by Klobuchar [18]. In Eq. (3), Re is the radius of the earth (6378 km), hmax is height (350 km), and θ is the elevation angle of satellite in radians.

$$VTEC = STEC * \cos\left\{\arcsin\left[\frac{R_e \cos\theta}{R_e + h_{\max}}\right]\right\} \tag{3}$$

3.3 Nakagami-m Distribution

Nakagami distributions are mostly used in electronics communication for modelling scattered radio frequency signals which arrive at the receiver through various paths. Due to this signal will have different fading characteristics. Rayleigh and Nakagami distributions are used to model dense scatters, and the Rician distributions are used to model fading with a stronger line of sight. If x has a Nakagami distribution with parameters μ and ω, then x^2 has a gamma distribution with shape parameter μ and scale parameter ω/μ (μ > ½ and ω > 0). Its probability density function (pdf) is given as

$$F(x; \mu, \omega) = 2\left\{\frac{\mu}{\omega}\right\}^{\mu} \frac{x^{(2\mu-1)} e^{\frac{-\mu x^2}{\omega}}}{\Gamma(\mu)} \tag{4}$$

3.4 Gaussian Distribution Function (GDF)

GDF or normal distribution is a very commonly used probability distribution function in statistics analysis approach for representing real valued random variables whose distributions are unknown. It is also known as bell curve because of the shape of the distribution curve. Here GDF is utilized to determine variations in the TEC and analyzing it. The normal distribution is given by Eq. (5).

$$y = f(x \mid \mu, \sigma) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \tag{5}$$

In Eq. (5), μ is the mean, σ is the standard deviation. The standard normal distribution is written as [ϕ(x)] sets μ to zero and σ to 1.

4 Results

The data received by IRNSS receiver on 3rd October 2015 is used as sample data for analysis. Calculation of parameters like TEC and VTEC is done using Eqs. (2) and (3). GDF and Nakagami distribution function of TEC data is also calculated. Figures 3, 4, 5, 6 and 7 is representing the GDF of TEC on Y axis and TEC in TEC units (TECU) on x axis for all PRN, and PRN 1 to PRN 4 respectively. It can be observed that TEC is varying from 5 TECU to 80 TECU. It can be seen that TEC is following GDF and the

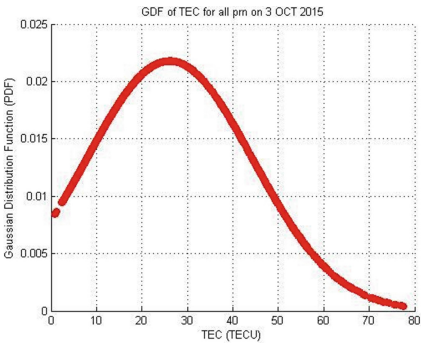


Fig. 3. GDF of TEC for all PRN from 1–4 on 3 Oct 2015.

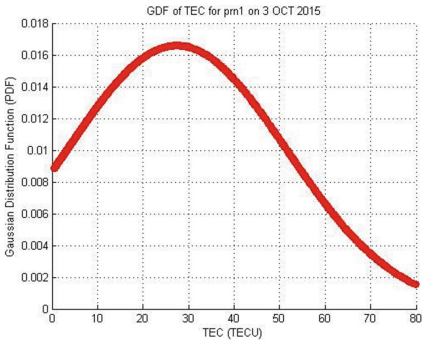


Fig. 4. GDF of TEC for PRN 1 on Oct 2015.

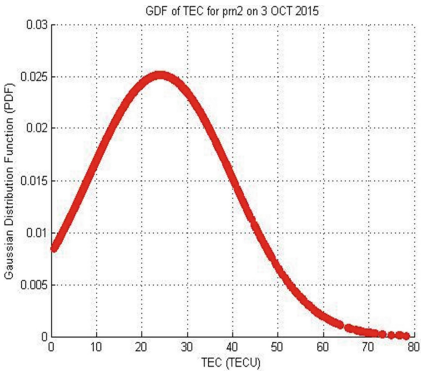


Fig. 5. GDF of TEC for PRN 2 on 3 Oct 2015.

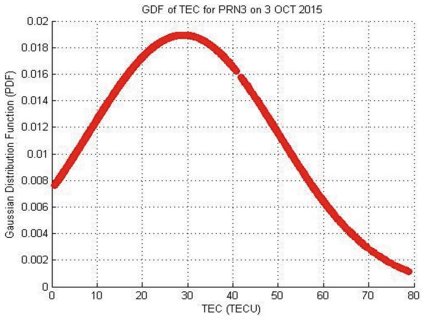


Fig. 6. GDF of TEC for PRN 3 on 3 Oct 2015.

shape of curve is bell shaped. The GDF values are varying from 0.008 to 0.025 for different values of TEC. Figures 8, 9, 10, 11 and 12 are representing comparison of GDF and Nakagami distribution function of TEC for all PRN and PRN 1–4 separately. TEC data are plotted using light blue colour bars as background of plot and GDF is indicated in red colour curve and Nakagami distribution function is indicated in blue colour on Y axis simultaneously for comparison purpose. From blue colour curve it is predicted that TEC variations are more following Nakagami distributions as compared to GDF.

Figures 13, 14, 15 and 16 are highlighting variations of Carrier to noise ratio (C/N) in decibel-Hertz and elevation angle in degrees for each PRN from PRN1 to PRN 4 for L5 and S band respectively. From Fig. 13 it seen that the C/N ratio is about 48 to 55 dB-Hz for L5 band and 43 to 47 dB-Hz for S band of PRN 1. And elevation angle

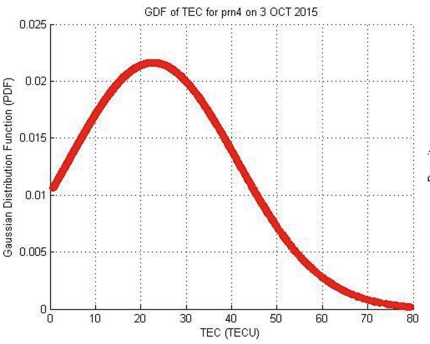


Fig. 7. GDF of TEC for PRN 4 on 3 Oct 2015.

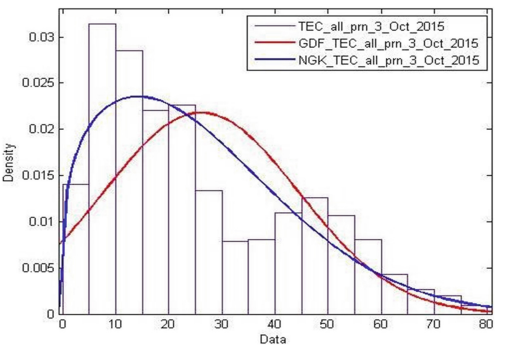


Fig. 8. Comparison of GDF and NGK of TEC for all PRN (1–4) on 3 Oct 2015. (Color figure online)

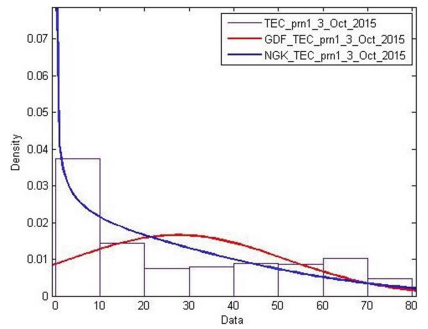


Fig. 9. Comparison of GDF and NGK of TEC for PRN 1 on 3 Oct 2015. (Color figure online)

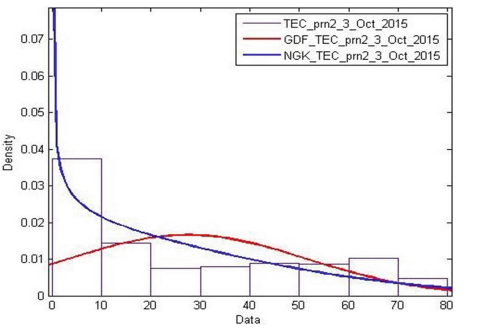


Fig. 10. Comparison of GDF and NGK of TEC for PRN 2 on 3 Oct 2015. (Color figure online)

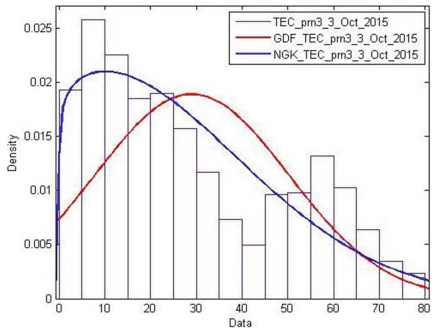


Fig. 11. Comparison of GDF and NGK of TEC for PRN 3 on 3 Oct 2015. (Color figure online)

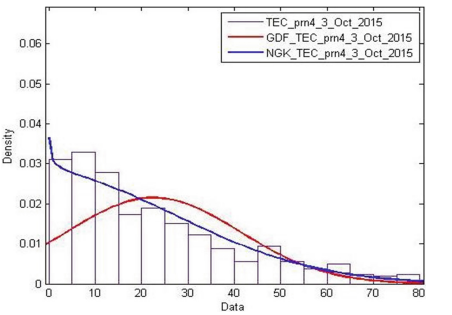


Fig. 12. Comparison of GDF and NGK of TEC for PRN 4 on 3 Oct 2015. (Color figure online)

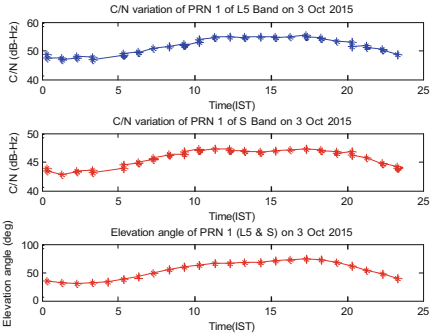


Fig. 13. Variation of C/N ratio and Elevation angle (L5 and S band) for PRN 1 on 3 Oct 2015

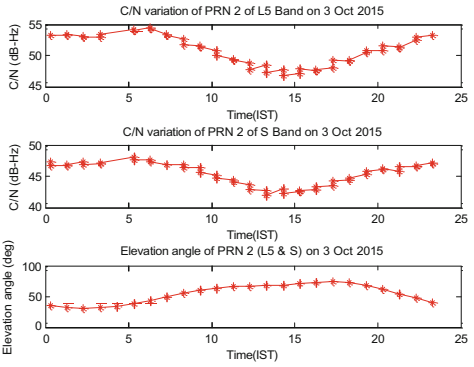


Fig. 14. Variation of C/N ratio and Elevation angle (L5 and S band) for PRN 2 on 3 Oct 2015

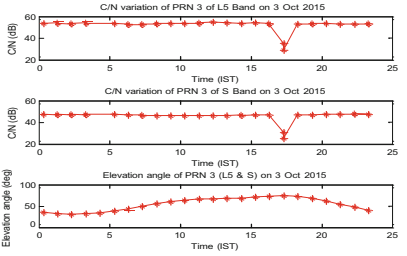


Fig. 15. Variation of C/N ratio and Elevation angle (L5 and S band) for PRN 3 on 3 Oct 2015

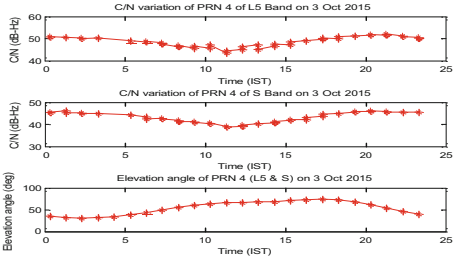


Fig. 16. Variation of C/N ratio and Elevation angle (L5 and S band) for PRN 4 on 3 Oct 2015

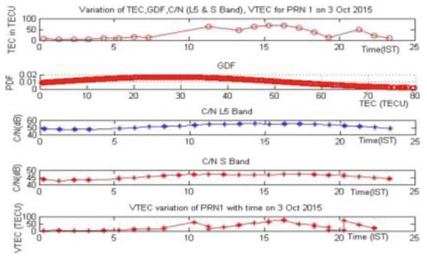


Fig. 17. Variation of TEC, GDF, C/N (L5 Band & S Band) for PRN 1 on 3 Oct 2015

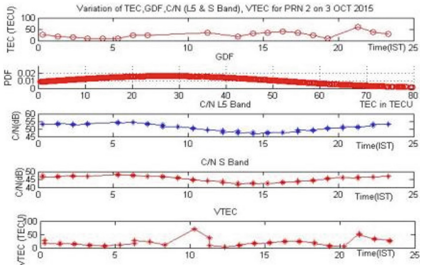


Fig. 18. Variation of TEC, GDF, C/N (L5 Band & S Band) for PRN 2 on 3 Oct 2015

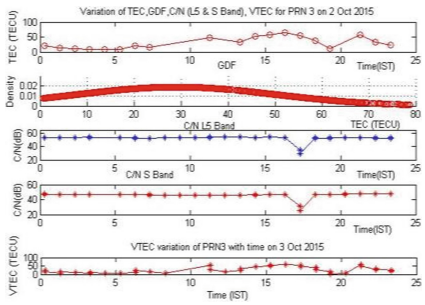


Fig. 19. Variation of TEC, GDF, C/N (L5 Band & S Band) for PRN 3 on 3 Oct 2015.

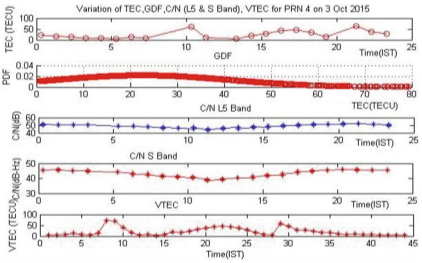


Fig. 20. Variation of TEC, GDF, C/N ratio (L5 Band & S Band) for PRN 4 on 3 Oct 2015.

of PRN1 is varying from 30° to 74°. From Fig. 14 it is seen that the C/N ratio is from 46 to 53 dB-Hz for L5 band and 41 to 48 dB-Hz for S band of PRN 2. And elevation angle of PRN 2 is varying from 30° to 74°. Figure 15 is indicating that C/N ratio for L5 band of PRN 3 from 29 to 54 dB-Hz and 25 to 47 dB-Hz for S band of PRN 3. Thus it is observed that for both L5 and S band the C/N ratio for PRN 3 is dropping around 25 to 30 dB-Hz around 17:30 IST. This can result in loss of lock of receiver with PRN 3 during this period. And elevation angle of PRN 3 is varying from 30° to 74°. Figure 16 is showing that C/N ratio for L5 band of PRN 4 is varying from 44 to 51 dB-Hz and 38 to 46 dB-Hz for S band of PRN 4. And elevation angle is varying from 30° to 74°. Figures 17, 18, 19 and 20 is showing variations in TEC in TECU on (Y-axis) versus IST local time (x-axis), GDF (Y-axis) of TEC data (x-axis), C/N ratio (Y-axis) in dB-Hz for L5 and S band with IST local time and VTEC in TECU versus IST local time for all PRN and PRN 1 to 4 respectively.

Thus the variations of various parameters like TEC, C/N ratio, Elevation angle and VTEC is carried out here for IRNSS signals captured on 3 Oct 2015.

5 Conclusions

Thus in this paper the variations of various parameters like carrier to noise power (C/N) ratio in dB-Hz, Elevation angle in degrees, Total Electron content (TEC) in TEC units (TECU) and Vertical TEC are studied for four satellites having PRN numbers from IRNSS 1A – IRNSS 1D (PRN 1–4). The Gaussian distribution function and Nakagami-m model distribution functions are also utilized for understanding behavioral pattern of TEC variations in this low latitude station and it is seen from the results that the TEC variations are more following the Nakagami distributions as compared to Gaussian distribution functions. The results can be further utilized for modelling ionospheric scintillation parameters in this region.

References

1. Kintner, P.M., Humphreys, T.E., Hinks, J.C.: GNSS and Ionospheric Scintillation-How to survive the next solar maximum, pp. 22–29 (2009). www.insidegnss.com
2. Saini, M., Gupta, U.: Indian GPS satellite navigation system: an overview. Int. J. Enhanc. Res. Manag. Comput. Appl. **3**(6), 32–37 (2014)
3. Senanayake, I.P.: Anticipated prospects and civilian applications of Indian satellite navigation services in Srilanka. Egypt. J. Remote Sens. Space Sci. **16**, 1–10 (2013)
4. Zaminpardaz, S., Teunissen, P.J.G., Nadarajah, N.: IRNSS stand-alone positioning: first results in Australia. J. Spat. Sci. **61**(1), 5–27 (2016)
5. Yashaswini, A.R., Reddy, P.S.N., Ramaiah, G.N.K.: Generation and implementation of IRNSS standard positioning signal. Eng. Sci. Technol. **19**, 1381–1389 (2016)
6. Sekhar, C.R., Srilatha, V.B.S., Dutt, S.I., Rao, S.: GDoP estimation using simulated annealing for GPS and IRNSS combined constellation. Eng. Sci. Technol. **19**, 1881–1886 (2016)
7. Thombre, S., Bhuiyah, M.Z.H., Soderholm, S., Jaakola, M.K., Ruotsaiaheh, L., Kuushiemi, H.: Tracking IRNSS satellites for multi-GNSS positioning in Finland. Inside GNSS **9**, 52–57 (2014)
8. Majithiya, P., Khatri, K., Hota, J.K.: Indian Regional Navigation Satellite System, correction parameters for timing group delays. Inside GNSS **6**, 40–46 (2011)
9. Manjula, T.R., Raju, G.: A comprehensive study of linear Kalman filter based tracking techniques under ionosphere scintillation. In: Global Colloquium in Recent Advancement and Effectual Researches in Engineering, Science and Technology (RAEREST 2016) (2016). Procedia Technol. 25, 427–434 (2016)
10. Manjunatha, R.N., Kiran, B.: Determination and preliminary analysis of position accuracy on IRNSS satellites. In: International Conference on Communication and Signal Processing, pp. 0765–0769, April 2016
11. Kumar, V., Hari, H.B., Pandiyan, R.: Real-time kinematic absolute and relative velocity estimation of geostationary satellites in formation using IRNSS observables. In: Third International Conference on Advances in Control and Optimization of Dynamical Systems, pp. 242–249, March 2014
12. Misra, P., Enge, P.: Global Positioning System, Signals, Measurements, and Performance. Ganga Jamuna Press, Lincoln (2001)
13. Van Dierendonck, A.J.: Eye on the ionosphere: measuring ionospheric scintillation effects from GPS signals. GPS Solut. **2**(4), 60–63 (1999)
14. Tanna, H.J., Karia, S.P., Pathak, K.N.: A study of L band scintillations during the initial phase of rising solar activity at an Indian low latitude station. Adv. Space Res. **52**, 412–421 (2013)
15. Sunda, S.: Simultaneous study of ionospheric total electron content and L-Band scintillations over the Indian equatorial Appleton anomaly stations. Ph.D. thesis, Department of Physics, Faculty of Engineering, Mohanlal Sukhadia University, September 2013
16. Parmar, S., Dalal, U., Pathak, K.N.: A study of diurnal variation of Ionospheric Scintillation effects on GPS signals at low latitude equatorial anomaly station, Surat, India. In: 2015 International Conference on Industrial Instrumentation and Control (ICIC), Pune, pp. 944–949 (2015)

17. Parmar, S., Dalal, U., Pathak, K.N.: Analysis of total electron content using GDF and Nakagami-m Distribution for Indian Regional Navigation Satellite system (IRNSS) signals at low latitude station, Surat, India, International Union of Radio Science, URSI-RCRS 2017, 3rd Regional Conference on Radio Science, Tirupati, India, March 2017
18. Klobuchar, J.: Design and characteristics of the GPS ionospheric time-delay algorithm for single frequency users. In: Proceedings of PLANS 1986 – Position Location and Navigation Symposium, Las Vegas, Nevada, pp. 280–286 (1986)

Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM) Model for Mobile Social Networks (MSN)

Zunnun Narmawala(✉)

Institute of Technology, Nirma University, Ahmedabad, India
zunnun80@gmail.com

Abstract. Peer-to-peer opportunistic communication between mobile devices carried by humans without using any infrastructure is largely unexploited. The encounter pattern of the devices depends on human mobility pattern which is governed by human social behaviour. Individuals belong to multiple communities. These social ties significantly affect humans' movement pattern. Traditional mobility models, such as Random Way Point (RWP) and Brownian Motion (BM), model device mobility as random. However, researchers have shown that human mobility is rarely random and such models do not provide a reliable analysis of network protocol performance. Various characteristics of human mobility are derived in the literature from mobility traces and social network theory. None of the mobility models in the literature incorporate all of them. In this paper, Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM) model is proposed incorporating all of these characteristics.

Keywords: Mobility model · Overlapping community structure
Opportunistic networks · Mobile Social Networks · ONE simulator

1 Introduction

In recent times, the growth of mobile devices (especially smartphones) is phenomenal. These devices support Bluetooth and Wifi connectivity. As these devices are carried by humans, their encounter patterns depend on human mobility patterns. Thus, knowledge of human movement behaviour and social structure can be exploited for efficient peer-to-peer communication [1,3] between these devices. As a result, this network paradigm is called as Mobile Social Network (MSN).

To analyze the performance of protocols which aim to exploit human movement behaviour through simulation, it is essential to design realistic mobility models which can mimic human mobility patterns as closely as possible. A number of experimental projects have been undertaken to collect encounter information of devices carried by humans [4,12]. These traces can be used in the

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 137–146, 2018.
https://doi.org/10.1007/978-3-319-73712-6_14

simulation to evaluate and analyze the performance of different protocols. While this approach generates realistic mobility patterns, its usefulness is limited as the performance of a protocol can be evaluated only for limited values of network parameters for which traces are available. Nonetheless, from analysis of these traces, various statistical properties of human mobility are derived [4,7,12,27]. Well-known and widely used mobility models such as Random Way Point (RWP) [14], Brownian Motion (BM) [8] etc. do not exhibit these properties. Further, movement of nodes is not independent. Nodes move as per the underlying overlapping community structure of humans who carry them. These mobility characteristics have a significant impact on the performance of forwarding strategy.

Community Aware Heterogeneous Human Mobility (CAHM) model [21] incorporates all these trace-based and social characteristics of human mobility. But, CAHM does not incorporate one important property of human mobility, i.e. locations that share many common users visiting them frequently tend to be located close to each other. In this paper, CAHM is improved by incorporating this property and this improved CAHM is called as Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM).

In the following Sect. 2, literature survey of existing mobility models for Mobile Social Networks (MSN) is presented. The proposed Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM) model is described in Sect. 3. Simulation results are discussed in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Literature Survey

To study characteristics of human mobility, many experimental studies at various universities (UCSD [18], Dartmouth [10], MIT [4], and University of Illinois [28]) and conferences (Infocom 2005 [12], Infocom 2006 [3], and SIGCOMM [24]) have been undertaken. In these experiments, humans participating in the experiment carry devices equipped with Wifi/Bluetooth and/or GPS sensor. These devices log encounter, location, and time information for a period of time.

From the analysis of these traces, various statistical properties of human mobility are derived which are as follows.

T.1 Aggregate inter-contact time follows power-law distribution with exponential cutoff [3,12].
T.2 Pause time follows truncated power-law distribution [27].
T.3 Humans visit nearby locations more frequently compared to far-away locations [7].
T.4 Humans have location preferences and they periodically re-appear at these locations [7].
T.5 Speed at which humans move increases with distance to be traveled [27].

2.1 Real-Trace Based Models

Real-trace based models try to capture features of individual's independent movement observed from mobility traces. Working Day Mobility (WDM)

model [5] and Time Variant Community (TVC) model [11] incorporate properties T.1 and T.4. Small World In Motion (SWIM) model [19] incorporates all properties T.1 to T.5. Self-similar Least Action Walk (SLAW) model [17] incorporates properties T.1, T.2, and T.3.

2.2 Social-Aware Models

Following are the main characteristics derived from the social network theory which affect human mobility.

S.1 Humans form communities based on their social relationships [22].
S.2 Humans belong to multiple communities and so, communities overlap [23].
S.3 Different individuals have different local popularity within a community and different global popularity in the social network [13].
S.4 Community size, the number of communities in which a node is a member and overlap size approximately follow power-law distribution where overlap size is defined as the number of individuals which are common in two communities [23].
S.5 Locations that share many common users visiting them frequently tend to be located close to each other [15].

Community-based Mobility Model (CMM) [20], Home-cell CMM (HCMM) [2], and N-body [31] models incorporate only S.1 of social network theory based properties. CMM and HCMM also incorporate some of the properties derived from mobility traces. But, these models do not incorporate properties S.2, S.3, S.4, and S.5 which are very important properties and have a significant impact on the performance of routing protocols. Social, sPatial, and Temporal mobility framework (SPoT) [15] is flexible and controllable mobility framework. But, it generates only contact traces and proposal in the paper for generating movement traces is preliminary. Further, it takes a social graph as an input instead of generating community structure synthetically. So, it lacks the flexibility of generating a large number of different social graphs for simulation. A detailed review of human mobility in opportunistic networks is available in [26].

Community Aware Heterogeneous Human Mobility (CAHM) model incorporates properties S.1 to S.4 derived from social network theory to generate community structure synthetically. CAHM is able to generate any number of overlapping community structures on its own based on input parameters. It does not take real life social network as an input, as requiring real life social network as an input restricts possible scenarios for which performance evaluation can be done. Further, it also does not use Social Network Models (SNM) such as Caveman model [29] to generate community structure, as these models are quite simplistic and do not take into account all social network theory based properties. It also incorporates all trace-based properties. However, it does not incorporate property S.5. So, Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM) model is proposed in this paper to incorporate property S.5. The summary of the comparison of different mobility models for MSN is presented in Table 1.

Table 1. Comparison of mobility models for MSN

Mobility model	T.1	T.2	T.3	T.4	T.5	S.1	S.2	S.3	S.4	S.5
SLAW [17]	✓	✓	✓							
WDM [5]	✓	✓	✓	✓						
TVC [11]	✓	✓	✓	✓						
SWIM [19]	✓	✓	✓	✓	✓					
N-body [31]	✓					✓				
CMM [20]	✓			✓		✓				
HCMM [2]	✓		✓	✓		✓				
HHW [30]	✓			✓		✓	✓	✓	✓	
CAHM [21]	✓	✓	✓	✓	✓	✓	✓	✓	✓	
P-CAHM (Proposed)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

As P-CAHM is based on CAHM, an overview of CAHM model is given in the following sub-section.

2.3 Overview of Community Aware Heterogeneous Human Mobility (CAHM) Model

In overlapping community structure, each individual n in the social network may belong to number of communities denoted as membership number Λ_n. Further, any two communities x and y may share $S^{ov}_{x,y}$ individuals, defined as overlap size between two communities. Let us denote size of community x as S^{com}_x and probability distribution functions of membership number, overlap size and community size as $\mathrm{P}(\Lambda)$, $\mathrm{P}(S^{ov})$ and $\mathrm{P}(S^{'com})$ respectively. Here, $S^{'com} = S^{com} - k$ to keep minimum community size equal to k where k is clique size. A k-clique is complete sub-graph of size k and k-clique community is union of all k-cliques that can be reached from one another through series of adjacent k-cliques where two k-cliques are adjacent if they share $k-1$ nodes [23]. Based on the analysis of a variety of social networks, Palla et al. [23] conclude that $\mathrm{P}(\Lambda)$, $\mathrm{P}(S^{ov})$ and $\mathrm{P}(S^{'com})$ approximately follow power-law distribution $\mathrm{P}(x) \sim x^{-\tau}$, with exponents $\tau = \Upsilon_\Lambda$, $\tau = \Upsilon_{Osize}$ and $\tau = \Upsilon_{Csize}$, respectively. Further, they report that values of Υ_Λ and Υ_{Osize} are not less than 2, and the value of Υ_{Csize} is between 1 and 1.6. These statistical properties are used to synthetically construct k-clique overlapping community structure.

P-CAHM model is composed of four components: (1) Establishing overlapping community structure, (2) Generating heterogeneous local degree, (3) Mapping communities into geographical zones, and (4) Driving individual motion. These components are explained in following four sub-sections.

Establishing *k*-clique Overlapping Community Structure. A day (or a week or any time duration) is divided into periods, and overlapping community structures are different in each of these periods but is same in the same period

of different days. Let us define nodes with membership number larger than 2, equal to 2 and equal to 1 as M-3 nodes, M-2 nodes, and M-1 nodes respectively. Community structure for each period is constructed as follows:

1. Generate nodes' membership numbers such that they follow $\mathrm{P}(\Lambda)$ with exponent Υ_{Λ}. Then, establish initial empty communities whose sizes S^{com} follow $\mathrm{P}(S^{'com})$ with exponent Υ_{Csize} such that $\sum_i \Lambda_i = \sum_j S_j{}^{\mathrm{com}}$.
2. Use all M-3 nodes to establish initial overlaps between pairs of communities.
3. Modify initial overlaps by allocating all M-2 nodes to communities such that overlaps' sizes follow $\mathrm{P}(S^{ov})$ with exponent Υ_{Osize}.
4. Allocate all M-1 nodes to unsaturated communities.

Generating Heterogeneous Local Degree. Local degree of a node within a community is defined as the number of neighbours of the node in the community. A node's local popularity depends on its local degree. Let $Local_i^n$ denote local degree of node n in its community i where $Local_i^n \geq k-1$ as per the definition of k-clique community. These values are generated such that they follow a power-law distribution with exponent Υ_{Local}.

Mapping Communities into Geographical Zones. To simulate n mobile nodes in a two-dimensional square plane, the model divides the plane into a grid of non-overlapping square cells. For each period, a community x with size S_x is associated with a zone composed of C_x adjacent cells. The location of a zone within the simulation plane is chosen randomly such that zones of different communities do not overlap. Each node n is randomly associated with $Local_i^n$ cells within the zone of its community i. Let μ_x be the average local degree and N_x be the number of nodes in community x. Let m be the community density index denoting denseness of a community. Then,

$$C_x = m \times \mu_x \times N_x \tag{1}$$

Driving Individual Motion. Initially, each node randomly selects one of its associated cells and then it is located at a random position inside that cell. To move, a node chooses an associated cell as next goal based on the distance it will have to travel with truncated power-law distribution $\mathrm{P}(D)$ with exponent Υ_D between the minimum distance and the maximum distance a node can travel.

As found in [27], speed increases with the increase in flight length because individuals use transportation to travel long distances instead of walking. They have also derived following relation between flight time (t) and flight length (l) from different mobility traces.

$$t = p \times l^{1-\eta}, 0 \leq \eta \leq 1 \tag{2}$$

From mobility traces, Rhee et al. [27] have proposed $p = 30.55$ and $\eta = 0.89$ when $l < 500\,\mathrm{m}$, and $p = 0.76$ and $\eta = 0.28$ when $l \geq 500\,\mathrm{m}$. CAHM uses this model to calculate speed at which a node should travel to next goal.

The overlapping community structure, corresponding associated zones and cells change at the start of the new period. When the period changes, after reaching its current goal, the node selects next goal inside one of its newly associated cells of the new period.

3 Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM) Model

In CAHM, the location of a zone associated with a community is selected randomly. But, locations of communities are not random. As shown in [15], locations that share many common users visiting them frequently tend to be located close to each other. So, CAHM is modified such that distances between communities are proportional to the number of common members of communities.

To decide the location of zones associated with communities, consider the network of communities as a graph where communities are nodes and two communities are connected by an edge if they have some common member nodes. Initially, zones are placed randomly in the simulation plane such that two zones do not overlap. Our goal is to place these zones such that distance between them is proportional to the number of common member nodes in corresponding communities.

Consider this as the n-body problem of physics. Two zones attract and repel each other with the force proportional to the number of common member nodes of the corresponding two communities. The pseudo-code is presented in the Algorithm 1. The algorithm is based on the one presented in [6] to draw a graph such that all vertices are placed at equal distance from each other. We need to place communities at distances which are proportional to the number of common member nodes of communities and instead of a point on the plane, a community requires an area on the plane.

In the algorithm, there are four steps in each iteration: calculate the effect of attractive forces on each community, then calculate the effect of repulsive forces, limit the total displacement by the 'temperature', and translate new positions of communities such that they are within simulation area. In using the 'temperature', the idea is to limit maximum displacement of a community to some maximum value, and this maximum value decreases over time. So, as the layout becomes better, the amount of adjustment becomes finer.

4 Simulation Results

P-CAHM model is implemented in ONE simulator [16]. It is a de facto simulator for Delay Tolerant Network (DTN) research. P-CAHM is simulated with the following scenario. There are 500 nodes in a simulation plane of 40 km × 40 km, divided into a grid of cells with size 252 m × 252 m each. The transmission range of each node is 40 m. The speed follows Eq. 2 and pause time is generated using power-law distribution with exponent 2 between 0 and 1000 s. 4-clique communities are generated, i.e. $k = 4$. Power-law exponents are set with $\Upsilon_\Lambda = 3$,

Algorithm 1. Algorithm to place communities based on number of common member nodes

$simulation\_area = maxX * maxY$
$G = (V, E)$ {Initial positions of communities V are random}
{k is the desired distance between mid-points of two communities and x is the current distance}
function $f_a(k, x) =$ begin return x^2/k end
function $f_r(k, x) =$ begin return k^2/x end
for $i = 1$ to $iterations$ **do**
  {Calculate repulsive forces}
  for v in V **do**
    {Each vertex has two vectors: .pos and .disp where .pos represents mid-point of a community}
    $v.disp = 0$
    for u in V **do**
      if $u \neq v$ **then**
        {Δ is the short hand for the difference vector between the positions of the two vertices}
        $\Delta = v.pos - u.pos$
        {r_v is the radius of community v, $tieStrength(u, v)$ represents number of common
        member nodes of u and v scaled between 0 and 1}
        $k = r_v + r_u + (1 - tieStrength(u, v))/(1 - avgTieStrength) *$
        $\sqrt{(maxX * maxY - totalCommunityArea)/|V|}$
        $v.disp = v.disp + (\Delta/|\Delta|) * f_r(k, |\Delta|)$
      end if
    end for
  end for
  {Calculate attractive forces}
  for e in E **do**
    {Each edge is an ordered pair of vertices $.v$ and $.u$}
    $\Delta = e.v.pos - e.u.pos$
    $k = r_v + r_u + (1 - tieStrength(u, v))/(1 - avgTieStrength) *$
    $\sqrt{(maxX * maxY - totalCommunityArea)/|V|}$
    $e.v.disp = e.v.disp - (\Delta/|\Delta|) * f_a(k, |\Delta|)$
    $e.u.disp = e.u.disp + (\Delta/|\Delta|) * f_a(k, |\Delta|)$
  end for
  {Limit the maximum displacement to the temperature t}
  for v in V **do**
    $v.pos = v.pos + (v.disp/|v.disp|) * min(v.disp, t)$
  end for
  for v in V **do**
    {Prevent from being displaced outside frame}
    $v.pos.x = translate(v.pos.x, min(.pos.x), max(.pos.x), r_v, maxX - r_v)$
    $v.pos.y = translate(v.pos.y, min(.pos.y), max(.pos.y), r_v, maxY - r_v)$
  end for
  {Reduce the temperature t as layout approaches better configuration}
  $t = cool(t)$
end for

$\Upsilon_{Osize} = 2$, $\Upsilon_{Csize} = 1.2$, $\Upsilon_{Local} = 2.4$, and flight length exponent $\Upsilon_D = 2$. All these values are in the range recommended for these exponents in the literature from mobility traces [22,23,27]. With a random seed, the model generates 14 communities with sizes 8, 121, 70, 6, 227, 7, 51, 91, 22, 3, 157, 3, 3, and 12. Because of space constraint, figures are not included. We run the simulation for 72,000 s.

To verify that in P-CAHM also, similar to CAHM, aggregate inter-contact time distribution is power-law with exponential cutoff, the simulation is done for two days. Simulation result shows that Complementary Cumulative Distribution Function (CCDF) of aggregate inter-contact times of P-CAHM follows power-law distribution with exponential cutoff which matches with the CCDF of aggregate inter-contact times of mobility traces [21].

To check the efficacy of our algorithm for the placement of communities proportional to the distances between them, Spearman's rank correlation coefficient (ρ) [25] is used. First of all, for initial random placement of communities, distances between communities are calculated and ordered list of initial distances is generated. ρ for this ordered list and the ordered list of tie strengths between communities comes out to be 0.19. Here, tie strengths between communities are number of common member nodes of communities scaled between 0 and 1. It shows that initially there is very weak correlation between distances and tie strengths. After the completion of the algorithm, the ρ comes out to be 0.52 which denotes a strong correlation between distances and tie strengths.

To check the effect of proximity property on the network performance in the given network scenario, 1/12 messages per second are generated in the network with the message size of 8 kB. In the steady state, with P-CAHM model, average message delivery delay and delivery ratio are 18000 s and 55% respectively. With CAHM model, they are 19029 s and 51%. In P-CAHM, as common member nodes need to travel less distances between communities, delivery delays of the messages they carry get reduced as compared to CAHM. As a consequence, less number of messages time out. So, the delivery ratio also improves.

5 Conclusion

To analyze the performance of routing protocols aiming to exploit human movement behaviour through simulation, it is essential to design realistic mobility models which can mimic human mobility patterns as closely as possible. Various characteristics of human mobility are derived from mobility traces and from social network theory in the literature. No existing mobility model, except CAHM, generates community structure synthetically incorporating all these characteristics and without using Social Network Models such as Caveman model. In this paper, Proximity and Community Aware Heterogeneous Human Mobility (P-CAHM) model is proposed with the following modification in CAHM: Instead of placing communities at random locations in the simulation plane, they are placed such that distances between them are proportional to the number of common member nodes of the communities. Simulation result

demonstrates that P-CAHM successfully establishes a strong correlation between distances among communities and number of common member nodes of communities. Also, CCDF of inter-contact times in P-CAHM follows power-law distribution as desired. Further, CAHM model under-reports network performance as compared to P-CAHM. The ONE simulator along with the P-CAHM mobility model can be downloaded from https://sites.google.com/a/nirmauni.ac.in/zunnun/.

References

1. Boldrini, C., Conti, M., Passarella, A.: Impact of social mobility on routing protocols for opportunistic networks. In: IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2007, pp. 1–6. IEEE (2007)
2. Boldrini, C., Passarella, A.: HCMM: modelling spatial and temporal properties of human mobility driven by users social relationships. Comput. Commun. **33**(9), 1056–1074 (2010)
3. Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., Scott, J.: Impact of human mobility on opportunistic forwarding algorithms. IEEE Trans. Mob. Comput. **6**(6), 606–620 (2007)
4. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Pers. Ubiquit. Comput. **10**(4), 255–268 (2006)
5. Ekman, F., Keränen, A., Karvo, J., Ott, J.: Working day movement model. In: Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models, pp. 33–40. ACM (2008)
6. Fruchterman, T.M., Reingold, E.M.: Graph drawing by force-directed placement. Softw. Pract. Exp. **21**(11), 1129–1164 (1991)
7. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
8. Groenevelt, R., Altman, E., Nain, P.: Relaying in mobile ad hoc networks: the Brownian motion mobility model. Wireless Netw. **12**(5), 561–571 (2006)
9. Han, B., Hui, P., Kumar, V., Marathe, M.V., Pei, G., Srinivasan, A.: Cellular traffic offloading through opportunistic communications: a case study. In: Proceedings of the 5th ACM Workshop on Challenged Networks, pp. 31–38. ACM (2010)
10. Henderson, T., Kotz, D., Abyzov, I.: The changing usage of a mature campus-wide wireless network. Comput. Netw. **52**(14), 2690–2712 (2008)
11. Hsu, W.J., Spyropoulos, T., Psounis, K., Helmy, A.: Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. IEEE/ACM Trans. Netw. **17**(5), 1564–1577 (2009)
12. Hui, P., Chaintreau, A., Scott, J., Gass, R., Crowcroft, J., Diot, C.: Pocket switched networks and human mobility in conference environments. In: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking, pp. 244–251. ACM (2005)
13. Hui, P., Crowcroft, J., Yoneki, E.: Bubble rap: social-based forwarding in delay-tolerant networks. IEEE Trans. Mob. Comput. **10**(11), 1576–1589 (2011)
14. Hyytiä, E., Koskinen, H., Lassila, P., Penttinen, A., Roszik, J., Virtamo, J.: Random waypoint model in wireless networks. In: Networks and Algorithms: Complexity in Physics and Computer Science, Helsinki (2005)

15. Karamshuk, D., Boldrini, C., Conti, M., Passarella, A.: SPoT: representing the social, spatial, and temporal dimensions of human mobility with a unifying framework. Pervasive Mob. Comput. **11**, 19–40 (2014)
16. Keränen, A., Ott, J., Kärkkäinen, T.: The one simulator for DTN protocol evaluation. In: Proceedings of the 2nd International Conference on Simulation Tools and Techniques, p. 55. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2009)
17. Lee, K., Hong, S., Kim, S.J., Rhee, I., Chong, S.: SLAW: a new mobility model for human walks. In: INFOCOM 2009, pp. 855–863. IEEE (2009)
18. McNett, M., Voelker, G.M.: Access and mobility of wireless PDA users. ACM SIGMOBILE Mob. Comput. Commun. Rev. **9**(2), 40–55 (2005)
19. Mei, A., Stefa, J.: SWIM: a simple model to generate small mobile worlds. In: INFOCOM 2009, pp. 2106–2113. IEEE (2009)
20. Musolesi, M., Mascolo, C.: Designing mobility models based on social network theory. ACM SIGMOBILE Mob. Comput. Commun. Rev. **11**(3), 59–70 (2007)
21. Narmawala, Z., Srivastava, S.: Community aware heterogeneous human mobility (CAHM): model and analysis. Pervasive Mob. Comput. **21**, 119–132 (2015)
22. Newman, M.E.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
23. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature **435**(7043), 814–818 (2005)
24. Pietilänen, A.K., Diot, C.: Dissemination in opportunistic social networks: the role of temporal communities. In: Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 165–174. ACM (2012)
25. Pirie, W.: Spearman rank correlation coefficient. In: Encyclopedia of Statistical Sciences (1988)
26. Pirozmand, P., Wu, G., Jedari, B., Xia, F.: Human mobility in opportunistic networks: characteristics, models and prediction methods. J. Netw. Comput. Appl. **42**, 45–58 (2014)
27. Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S.J., Chong, S.: On the levy-walk nature of human mobility. IEEE/ACM Trans. Networking (TON) **19**(3), 630–643 (2011)
28. Vu, L., Do, Q., Nahrstedt, K.: Jyotish: constructive approach for context predictions of people movement from joint Wifi/Bluetooth trace. Pervasive Mob. Comput. **7**(6), 690–704 (2011)
29. Watts, D.J.: Small Worlds: The Dynamics of Networks Between Order and Randomness. Princeton University Press, Princeton (1999)
30. Yang, S., Yang, X., Zhang, C., Spyrou, E.: Using social network theory for modeling human mobility. IEEE Network **24**(5), 6–13 (2010)
31. Zhao, C., Sichitiu, M.L., Rhee, I.: N-body: a social mobility model with support for larger populations. Ad Hoc Netw. **25**, 185–196 (2015)

Measuring the Effect of Music Therapy on Voiced Speech Signal

Pradeep Tiwari[1(✉)], Utkarsh V. Rane[1], and A. D. Darji[2]

[1] MPSTME, NMIMS University, Mumbai, India
pradeep.tiwari@nmims.edu, raneutkarsh@gmail.com
[2] SVNIT, Surat, India
addarji@gmail.com

Abstract. With the rapid development in the field of speech processing, the human speech is being analyzed from different perspectives. Now-a-days impact of external factors like music on speech are also being studied by the researchers. It is widely accepted fact that the music plays important role in refreshing the mood when we see most of the people listening to the music in train or bus to get rid of boredom. This paper deals with the relation between music & its effect on human speech based on the fact that brain (cerebrum) has control over vocal tract (speech). It is also observed that the people work efficiently while listening music to increase their alertness & concentration. By studying voice samples of fatigued persons (physically or mentally fatigued) of different age-groups, it has been observed that listening to music reduces considerably the average mean & the average standard deviation feature of the speech waveform. It has also been observed that average energy of the speech waveform gets reduced & its zero crossing rate (ZCR) gets increased.

Keywords: Music therapy · Speech Feature Extraction · Stress

1 Introduction

Stress is called the lifestyle disease in today's world that not only limits individual's capabilities, interest levels and mood but also causes physical and mental health problems as in [1]. Music therapy is getting more and more reputation nationally and internationally since it is painless, no side effects, and low cost treatment for depressed patients in [2]. Background music can also help enhancing the efficiency of individuals who work with their hands as it increases their alertness and concentration in [3]. Mental condition plays an important role in the course of recovery and affect the efficiency of administered medicines in the process of disease and cure and should be taken into consideration during diagnosis and treatment in [4].

Speech signal is composed of a sequence of sounds which are produced as a result of acoustic excitation of the vocal tract when air is expelled from the lungs as shown in Fig. 1. The first step of speech production is when the speaker formulating the message in mind which he intends to transmit to the listener via medium.

The message is then converted into language code with the help of set of phoneme sequences corresponding to the sounds that make up the words. The prosody which

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 147–156, 2018.
https://doi.org/10.1007/978-3-319-73712-6_15

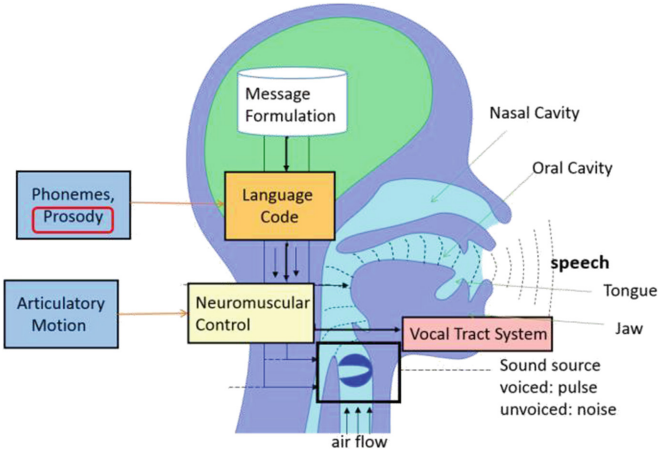


Fig. 1. Process of speech production

defines the stress is also added at this step in accordance with duration, loudness of the sounds & also the pitch associated with it as in [5].

Stressed Speech is defined as the speech produced under any condition that causes the speaker to vary the speech production from the neutral condition as in [6]. If a speaker is in a 'quiet room' with no task responsibilities, then the speech produced is considered neutral. Stress can be classified as (a) Emotionally induced stress: Speech produced under change in the emotional or psychological state of the speaker such as angry speech, sad speech, happy speech etc. (b) External Environmentally induced stress such as Lombard Speech (c) Pathological Stressed Speech such as Cold affected speech, Old age Speech. In this paper, emotionally and External environmentally induced stressed speech are considered.

Different subjects with stress were selected. Speech signal are acquired from each subjects before and after they were subjected to listening music. The recorded speech signal is sampled at sampling frequency of 44100 Hz and 16 bits per sample. The speech feature selection is an important problem in stress identification. Speech features such as average Energy, average mean, average Standard Deviation, average Zero Crossing Rate (ZCR) are obtained for the acquired speech signal from the subjects before and after hearing music. It was clearly seen that average energy, average mean & average standard deviation is found decreasing after hearing the music. Also it is observed that ZCR is increasing after listening the music.

2 Related Works

There have been considerable research works in the field of Neuropsychology and speech processing since past two decades. The assessment carried in [7] indicates that the relaxation and concentration improves using Alpha music, which influence the alpha and beta rhythms significantly. When subjects faced Alpha music, they felt

considerable reduction in fatigue/stress along with the increase in the physical relaxation. Thus it is evident that music affects human brain & relaxes it during fatigue. Another kind of work showing effect of music on human body & mind where subjects were made to listen to particular music for particular span of time & their heart rate variability was measured using ECG machine. It has been observed that the single cluster formed by the volume of the Point Care Plots of Spherical Coordinate is reduced remarkably in the data acquired during music state as compared to the data of pre music state & also the amount of reduction is not the same for all the subjects. This proves that music has some definite effect on human physiology [8]. Some investigations of researchers are even showing improvement in the typewriting work performance due to impressions of music by measuring some bio-signals to monitor participant's condition. Furthermore, the music impression causes activation in saliva amylase which decreases fatigue/stress as in [9]. A study is presented in [10] depicts that when a subject is exposed to live violin music performance, its brain induces theta, alpha and beta brainwaves to get balanced. Another study shows that to relieve users from depression, an electroencephalogram (EEG) based music therapy system was used to identify the user and to measure the degree of depression giving results which gives conclusion that the EEG approach is user-based approach for preventing depression [11]. A number of studies have been done on understanding the relation between music & brain & one such suggests that injury to brain can drastically impair musical activities except leaving intellectual and linguistic abilities. This research indicates that music cognition is not affected entirely, but in particular abilities [12]. Another work carried by a researcher presents that music enhances spatial-temporal reasoning when the data collected from College students & Preschool children. It also showed that the effect is more if the subjects are exposed to longer duration of time [13]. In [13] a structured neural model has been analyzed, describing a certain kind of relationship between music and spatial-temporal reasoning. The study presented in [14] has developed relationship between music, and subject's emotion and concentration with the help of EEG device. In [15] EEG device is used to check the emotional responses while listening music with the help of low cost cloud based architecture. The functioning of the human brain while listening to music is studied in [16] with the help of Natural stimulus functional magnetic resonance imaging (N-fMRI). The observation derived in this experimentation shows that music like classical music, pop etc. affects the attention and emotions of human brain [16].

3 Process Description

This section describes the block diagram of the project (Fig. 2).

3.1 Data Acquisition

This is the first step of the project. In this speech signal of sampling frequency 44.1 kHz & 16 bits/s using microphone has been acquired. The microphone of cellphone HTC Desire 620G is used for recording voices. The vowels a, e, i, o, u are considered to be spoken by each speaker as they are voiced sounds, hence these vowels

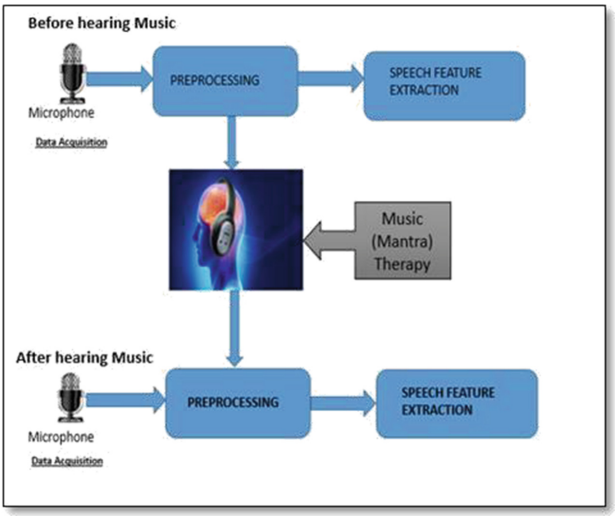


Fig. 2. Block diagram

contain high energy & maintains uniformity in the sentence spoken by each speaker. Ten voice samples of ten subjects, both male & female of various age groups are taken before hearing the music & then again after hearing the music. In this the subjects are made to hear sweet & soft romantic music. The phone's speakers used have following specifications: Stereo speakers of HiFi edition with one at the top of the phone acting as tweeter & other at the bottom acting as woofer & working together to offer experience of high quality surround effect with Dolby Audio 4. All the ten subjects are from Mumbai, India region speaking typical Indian accent English. The recording time for each sample was around 3–6 s. Mainly the age-groups are youngsters like between ages 18 to 26 of which the three voice samples of male & female each are acquired. Two voice samples of one male of age 64 & one female with the age of 62 are taken. Finally, the two voice samples of middle-aged group around 40 to 50 years of age are taken which are both females. So different aged voices are used to make project more extensive & authentic.

3.2 Data Preprocessing

The acquired data of ten subjects which is in '.wav' format is plotted as shown in Fig. 3. The waveforms of the speech sample before hearing and after hearing music is analyzed which reflect the fact that there is variation in the speech produced on hearing music. The five vowels were clipped separately for both the cases (i.e. before hearing the music & after hearing the music). Further preprocessing is done for '.wav' files. Preprocessing includes normalization & pre-emphasis. Direct current offset (de-offset) carries no useful information rather it can carry disturbing information. Removal of de-offset is called normalization. The statistical normalization which is widely given by,

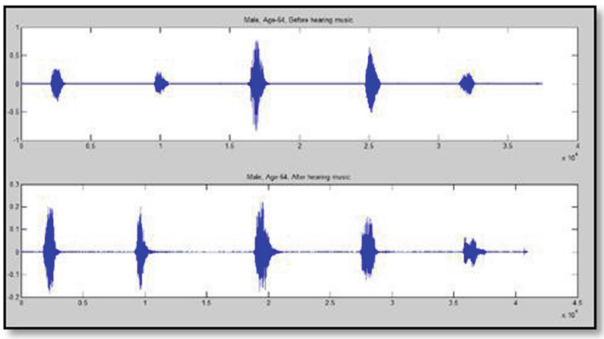


Fig. 3. Speech sample (a, e, i, o, u) waveform before and after hearing music

$$S_n = \frac{S - mean(S)}{Variance(S)} \tag{1}$$

The higher frequency components of speech signals are suppressed while speech production. Pre-emphasis increases the magnitude of the higher frequencies with respect to the magnitude of the lower frequencies. A simple first-order high pass, FIR filter is generally used for pre-emphasis as given below:

$$H(z) = 1 - kz^{-1} \quad \text{where k} \in [0.9, 1] \tag{2}$$

Next step is Speech Feature Extraction to identify the changes in each vowel spoken in both the cases.

3.3 Speech Feature Extraction

This is the next step in this research paper. In this step the features considered for analysis are Short time Energy, Zero Crossing Rate (ZCR) and Statistical features (Mean, Standard Deviation).

3.3.1 Short-Time Energy

It has been watched that the amplitude of the speech signal fluctuates considerably with time. Specifically, the sufficiency of the unvoiced portions is by and large lower than the amplitude of voiced portions. The short-time energy of speech signal gives a helpful portrayal that mirrors these amplitude fluctuations. As a rule, it can be characterized the short-time energy as in [17] as, this expression can be written as (Fig. 4),

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \tag{3}$$

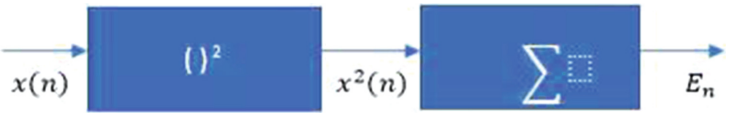


Fig. 4. Block diagram representation of the short-time energy [17]

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) \quad where\, h(n) = w^2(n) \tag{4}$$

The real noteworthiness is that it gives a premise for distinguishing voiced speech segments from unvoiced speech segments. Also, for very high quality speech (high signal-to-noise ratio), the energy can be used to distinguish speech from silence.

3.3.2 Short-Time Average ZCR

With regards to discrete time signals, a zero-crossing is said to happen if successive specimens have distinctive algebraic signs as in [17]. The rate at which zero-crossings happen is a basic measure of the frequency content of a signal. This is especially valid for narrowband signals. For example, a sinusoidal signal of frequency F_0 sampled at a rate F_s, has F_s/F_0 samples per cycle of the sine wave. Each cycle has two zero-crossings so that the long-time average rate of zero-crossings is

$$Z = \frac{2F_0}{F_s} crossing/sample \tag{5}$$

Thus, the average zero-crossing rate gives a reasonable way to estimate the frequency of a sine wave. Speech gives rough estimates of spectral properties can be acquired utilizing a representation of zero-crossing rate for the speech. A suitable definition is (Fig. 5),

$$Z_n = \sum_{-\infty}^{w} |\mathrm{sgn}[x(m)] - \mathrm{sgn}[x(m-1)]|w(n-m) \tag{6}$$

Where,

$$sgn[x(n)] = 1 \quad x(n) \geq 0$$
$$= -1 \quad x(n) < 0$$

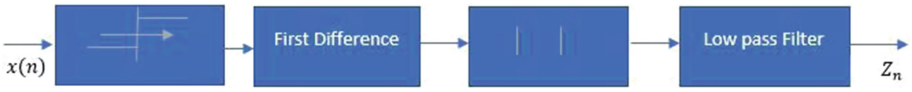


Fig. 5. Block diagram representation of short-time average zero-crossings [17]

And

$$w(n) = \frac{1}{2N} \quad 0 \le n \le N-1 \\ = 0 \qquad \text{otherwise} \tag{7}$$

Zero Crossing Rate (ZCR) represents the frequency content of a signal, since it checks number of times the amplitude of the speech signals passes through a value of zero in a given time interval/frame. The interpretation of average ZCR is less exact since speech or voice signals are broadband signals, however it can be generally estimated.

3.3.3 Statistical Features

The mean and standard deviation are the two prominent statistical features explored in this paper. The following equation gives Mean value for every vowel:

$$M = \sum_{i=1}^{n} \frac{x_i}{n} \tag{8}$$

The Standard Deviation value for every vowel is given by following equation:

$$S = \sqrt{\sum_{i=1}^{n} \frac{(x_i - M)}{n-1}}. \tag{9}$$

4 Implementation and Results

The core idea of undertaking this project is to identify & analyze the impact of music on the speech of person under fatigue. Both the types of fatigue are considered: mental & physical. The acquired speech samples generated for this project are vowels a, e, i, o & u. The ten samples of each five vowels are considered for before hearing and after hearing the music. The energy of each vowel is calculated as shown in the Fig. 6.

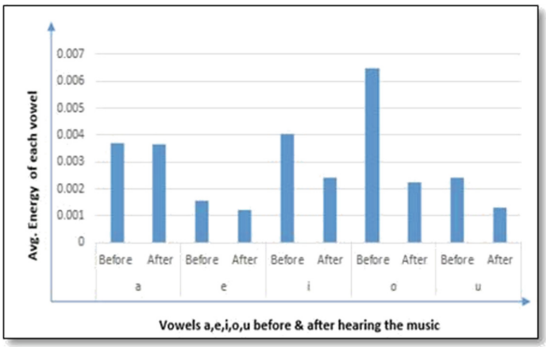


Fig. 6. Average Energy of vowels a, e, i, o, u before & after hearing music

As it can be seen from Fig. 6 that average energy values of each vowel are higher before hearing the music as compared to the value of average energy values after hearing the music. Similarly, it is depicted in the Fig. 7 that average mean values of each vowel are more before hearing the music than after hearing the music.

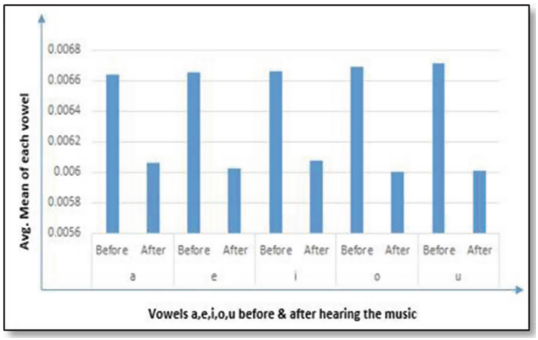


Fig. 7. Average Mean of vowels a, e, i, o, u before & after hearing the music

Also from the Fig. 8, it can be observed that average standard deviation values before hearing the music for each vowel is more than the values for vowels after hearing the music.

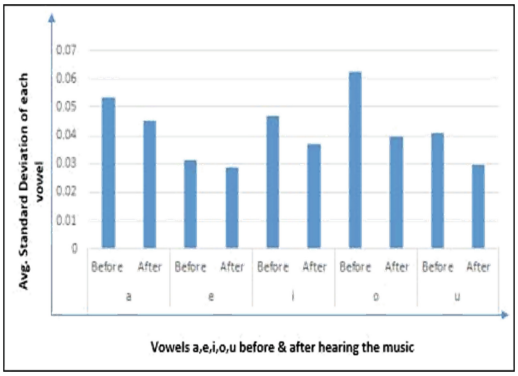


Fig. 8. Average Standard Deviation of vowels a, e, i, o, u before & after hearing the music

There is a difference which can be seen in the Fig. 9 for ZCR feature. The ZCR values for each vowel after hearing the music are more compared to the values before hearing the music. Though the values for vowel O are showing a kind of deflection from the general pattern but it can be generalized that the ZCR values are generally more after hearing the music than the ZCR values before hearing the music.

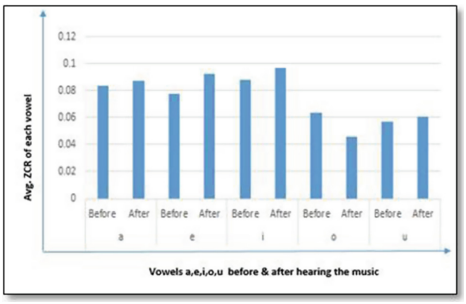


Fig. 9. Average Zero Crossing Rate of vowels a, e, i, o, u before & after hearing the music

5 Conclusion

In today's world, due to the competition and fast paced life, students, corporates and family often experience fatigue or stress which reduces their expected performance. This paper clearly gives an evidence that the speech is connected to & is in the control of brain by means of prosody. Furthermore, the paper concludes that the music affects the speech features, hence the stress level of the subject. The future scope of this project is that the statistical features of the speech & the facial features can be considered for better analysis of the impact of music on the speech.

References

1. Vandyke, D.: Depression detection & emotion classification via data-driven glottal waveforms. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp. 642–647. IEEE (2013)
2. Zhou, P., Lin, D., He, W., Li, G., Shang, K.: Influence of musicotherapy on mental status and cognitional function of patient with depression disease. In: 2009 2nd International Conference on Biomedical Engineering and Informatics, pp. 1–4. IEEE (2009)
3. Restak, R.: Mozart's Brain and Fighter Pilot. Crown Publications, New York (2003)
4. Cornelia, B., Richardson-Boedler, C.: Applying Bach Flower Therapy to the Healing Profession of Homoeopathy. B. Jain Publishers, New Delhi (2003)
5. Rabinar, L., Juang, B.H., Yeganarayana, B.H.: Fundamental of Speech Recognition. Pearson (2010). Second Impression
6. Ramamohan, S., Dandapat, S.: Sinusoidal model-based analysis and classification of stressed speech. IEEE Trans. Audio Speech Lang. Process. **14**(3), 737–746 (2006)
7. Vijayalakshmi, K., Sridhar, S., Khanwani, P.: Estimation of effects of alpha music on EEG components by time and frequency domain analysis. In: 2010 International Conference on Computer and Communication Engineering (ICCCE), pp. 1–5. IEEE (2010)
8. Das, M., Jana, T., Dutta, P., Banerjee, R., Dey, A., Bhattacharya, D.K., Kanjilal, M.R.: Study the effect of music on HRV signal using 3D Poincare plot in spherical co-ordinates-a signal processing approach. In: 2015 International Conference on Communications and Signal Processing (ICCSP), pp. 1011–1015. IEEE (2015)

9. Iwaki, M., Nakano, K.: Typewriting performance affected by music impression in working environment. In: 2013 Proceedings of SICE Annual Conference (SICE), pp. 1539–1543. IEEE (2013)
10. Hassan, H., Murat, Z.H., Ross, V., Buniyamin, N.: A preliminary study on the effects of music on human brainwaves. In: 2012 International Conference on Control, Automation and Information Sciences (ICCAIS), pp. 176–180. IEEE (2012)
11. Peng, H., Hu, B., Liu, Q., Dong, Q., Zhao, Q., Moore, P.: User-centered depression prevention: an EEG approach to pervasive healthcare. In: 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, pp. 325–330. IEEE (2011)
12. Peretz, I., Hébert, S.: Music processing after brain damage: the case of rhythm without melody. In: Steinberg, R. (ed.) Music and the Mind Machine, pp. 127–137. Springer, Heidelberg (1995). https://doi.org/10.1007/978-3-642-79327-1_13
13. Shaw, G.L.: Computation by symmetry operations in a structured neural model of the brain: music and abstract reasoning. In: Cabrera, B., Gutfreund, H., Kresin, V. (eds.) From High-Temperature Superconductivity to Microminiature Refrigeration, pp. 287–311. Springer, New York (1996). https://doi.org/10.1007/978-1-4613-0411-1_25
14. Sourina, O., Kulish, V.V., Sourin, A.: Novel tools for quantification of brain responses to music stimuli. In: Lim, C.T., Goh, J.C.H. (eds.) 13th International Conference on Biomedical Engineering, pp. 411–414. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-92841-6_101
15. Guo, Y., Wu, C., Peteiro-Barral, D.: An EEG-based brain informatics application for enhancing music experience. In: Zanzotto, F.M., Tsumoto, S., Taatgen, N., Yao, Y. (eds.) International Conference on Brain Informatics, pp. 265–276. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35139-6_25
16. Fang, J., Xintao, H., Han, J., Jiang, X., Zhu, D., Guo, L., Liu, T.: Data-driven analysis of functional brain interactions during free listening to music and speech. Brain Imaging Behav. **9**(2), 162–177 (2015)
17. Rabiner, L.R., Schafer, R.W.: Digital Processing of Speech Signals. Prentice Hall, Englewood Cliffs (1978)

Ensuring Database and Location Transparency in Multiple Heterogeneous Distributed Databases

Shefali Naik(✉)

School of Computer Studies, Ahmedabad University, Ahmedabad, India
shefali.naik@ahduni.edu.in

Abstract. Challenges and issues of distributed database are well known. One of the challenges is obtaining database transparency. In distributed database, the physical database is distributed across heterogeneous database systems. There are several methods of data access from distributed database, still it involves complicated implementation of these methods. Oracle DBMS provides heterogeneous gateway service to connect oracle and non-oracle databases. The data access and distributed transaction execution is made very easy through this gateway. In this paper, the procedure to connect four heterogeneous databases namely PostgreSQL, MySQL, Oracle and MS Access is described. The successful implementation of data access from multiple heterogeneous databases provides very easy and efficient access of data in Oracle using its native language/commands. Oracle user will neither have to bother about architecture nor commands of the remote databases from where data is accessed. Even user need not have to worry about the location of database. Once all the databases are connected through heterogeneous gateway, user will be able to access data and process distributed transactions efficiently from Oracle. The paper covers the whole process.

Keywords: Heterogeneous distributed database · Database transparency
Location transparency · Distributed transactions · Heterogeneous gateway

1 Introduction

In Distributed Database [2, 3], the physical database is distributed across multiple sites. One of the architecture of distributed database management system is multiple databases [3, 7]. There are many types of distributed database architectures. One of them is heterogeneous multiple distributed database [1, 3, 7]. In multiple heterogeneous distributed database environment, autonomous multiple databases residing on different locations [3, 7] are connected together to access data of each other. While connecting and accessing data, it should provide transparency at various levels to the users. The different levels of transparency are: Location Transparency, Replication Transparency, Fragmentation Transparency, Database Transparency, Operating System Transparency, Network Transparency, etc. [3–5]. To achieve location and database transparency,

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 157–163, 2018.
https://doi.org/10.1007/978-3-319-73712-6_16

Oracle provides heterogeneous gateway service [6]. The meaning and implementation of these two transparencies is explained in the following section.

2 Location and Database Transparency

In distributed database, with location transparency [3], user can access the data without knowledge of location of data. i.e., user doesn't have to specify the location of data in the commands. Similarly, with database transparency [3], user will be able to access data from many databases without knowledge of commands of all the databases. Figure 1 shows the multiple heterogeneous database architecture which connects PostgreSQL, MySQL, MS Access and Oracle database using Oracle's Heterogeneous Gateway Service.

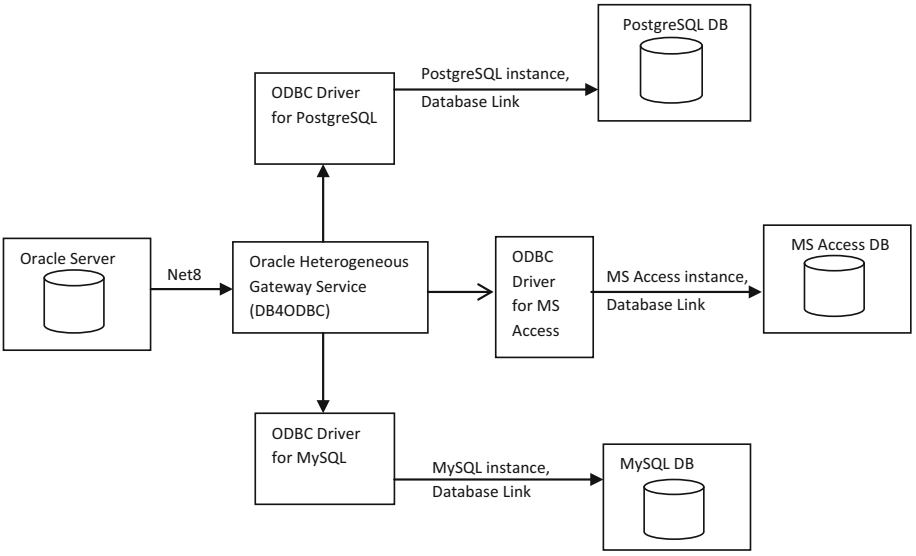


Fig. 1. Multiple heterogeneous database architecture

The implementation steps of accessing data from multiple heterogeneous databases [7] are explained in Sect. 3. The process of establishing connection between oracle and non-oracle database require deep knowledge of databases.

3 Implementation

3.1 Steps to Connect Oracle and Non-Oracle Databases

To access data [9] from multiple heterogeneous databases, the respective DBMS software along with its ODBC drivers must be installed. Oracle heterogeneous service

is also required which could be installed from oracle technology network. The client library of oracle is also needed. The sequence of flow given below should be followed to connect and access data from non-oracle database to oracle [6].

Step-1. Install Oracle Server, Oracle Client Library and Oracle Heterogeneous Gateway Service from oracle technology network.

Step-2. Install PostgreSQL, MySQL and MS Access databases with respective ODBC drivers.

Step-3. Create users in PostgreSQL and MySQL databases. Grant required privileges to the users. Create tables in specific user's account. Insert data in these tables.

Step-4. Open ODBC data source and create three system data source names for each postgreSQL, MySQL and MS Access databases. Test connections.

Step-5. After successful connections, create three initialization parameter files for each of the system DSN with name "init<DSN>.ora". Save this file in the folder where oracle home is installed. This file will contain name of the database instance, which is system DSN name.

Step-6. Open listener.ora and tnsnames.ora files from the folder where oracle home is installed. Modify both the files with required entries and save.

Step-7. Start listener service from command prompt. Check connection of each of the database instance using "tnsping <DSN>" command.

Step-8. Create database links in oracle client to access data from non-oracle database.

Step-9. Access data from non-oracle database using query language of oracle.

3.2 Flowchart of Oracle and MySQL Connection

The flowchart of whole process of oracle with MySQL connection and data access is given in Fig. 2. The process of connecting different heterogeneous databases could be simplified by writing a program which will automate the steps given in Fig. 2.

4 Achieving Location and Database Transparency

To access the data from heterogeneous databases [9] using the method described in Sects. 2 and 3, there is no need to learn syntax of programming language of all the databases. It could be done only by writing the database link name [9] with table name of a specific database. Thus, database transparency is achieved. The database which is stored on different host, there is no need to specify hostname or address of site anywhere in the commands. Hence, achieving the location transparency. Few examples of SQL commands are given in Fig. 3 and stored procedures are given in Fig. 4.

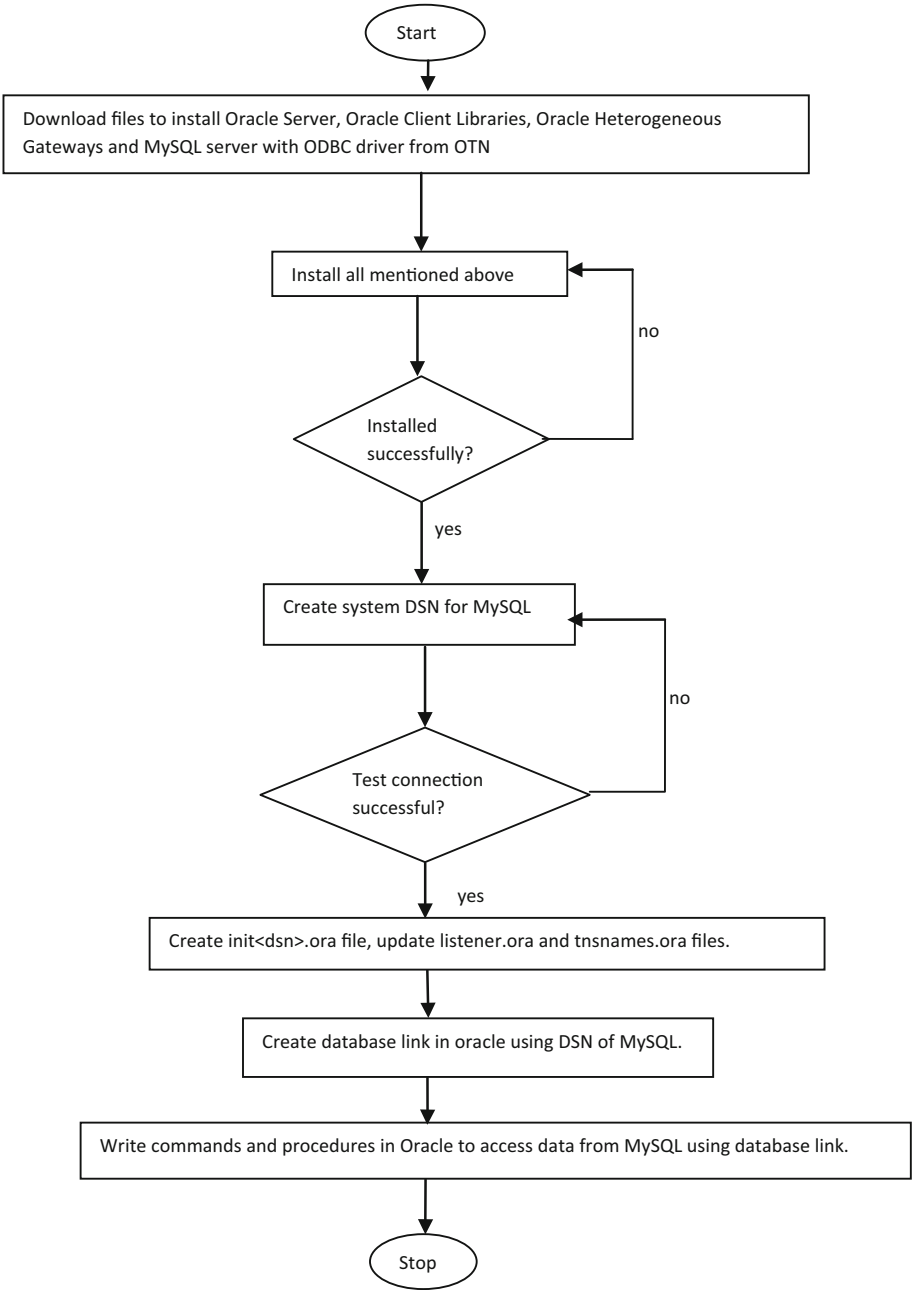


Fig. 2. Flowchart of connection of oracle and non-oracle databases

- Commands written in MySQL

```
mysql> use test;
mysql> create table emp(empno int, empname varchar(30));
mysql> insert into emp values(101,'Shefali');
mysql> create user 'shefali'@'localhost' identified by
'shefali';
mysql> grant all on *.* TO 'sys'@'%'  IDENTIFIED BY
'Shefali';
```

- Commands written in Oracle's SQL Plus

```
SQL>create public database link toms connect to "shefali"
identified by "shefali" using 'MS';

SQL> select * from "emp"@toms;

SQL> select e."eno" from "emp"@toms e;

SQL> select e."eno"||'   '||e."enmae"||'
'||s."sal_date"||'    '||s."salary"
  2  from "emp"@topg e, salary@toac s
  3  where e."eno"=s."empno";
```

Fig. 3. Commands written in oracle client to access data from heterogeneous databases

By creating synonyms [10], the global name could be defined for database link. The global name may be used instead of database link. For ex., "employee" synonym could be created using the following command. Then, "emp"@topg database link could be replaced with "employee".

```
CREATE SYNONYM employee FOR "emp"@topg;
```

```
Create or replace procedure disp as
              cursor c_emp is select * from
"public"."emp"@toms;
              r_emp c_emp%rowtype;
             cursor c_kg is select * from salary@toac;
             r_kg c_kg%rowtype;
begin
            dbms_output.put_line('Data from PostgreSQL…..');
             for r_emp in c_emp loop
                       dbms_output.put_line('Emp Name
==>'||r_emp. "enmae");
            end loop;

            dbms_output.put_line('Data from MS Access…..');
             for r_kg in c_kg loop
                       dbms_output.put_line('Emp No
==>'||r_kg. "empno");
            end loop;
end;
/
```

Fig. 4. Stored procedure written in oracle client to process data of heterogeneous databases

5 Conclusion and Future Work

Data access and process from multiple heterogeneous databases [9] is done very effectively using oracle heterogeneous service. The task is very complicated, which requires in depth knowledge of oracle database administration. Even it is very challenging for database administrators. Therefore, this whole process of creating multiple heterogeneous database environments should be automatic. It may be possible that the whole process could not be automatic, but it could be semi-automatic. The model and implementation of automation of this process could be done in future. It will be very useful if such type of model is developed which ease the process of creation of multiple heterogeneous database environment. Furthermore, work could be done to improve fragmentation, replication and allocation transparencies [3, 4] which will improve performance of distributed concurrent transaction [8] execution in multiple heterogeneous distributed database.

References

1. Ferrier, A., Stangret, C.: Heterogeneity in the distributed database management system SIRIUS-DELTA. In: VLDB (1982)
2. Naik, S.: Concepts of Database Management System. Dorling Kindersley, New Delhi (2014)
3. Özsu, M.T., Valduriez, P.: Principles of Distributed Database Systems. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-8834-8

4. Adiba, M.E., et al.: Issues in distributed data base management systems: a technical overview. In: Proceedings of the Fourth International Conference On Very Large Data Bases, vol. 4. VLDB Endowment (1978)
5. Shefali, N., Samrat, K.: Revisited performance issues in concurrent transaction execution in distributed database management system. Int. J. Curr. Eng. Sci. Res. **2**(4), 23–26 (2015). ISSN (Print): 2393-8374, (Online): 2394-0697
6. Oracle Database Heterogeneous Connectivity User's Guide. https://docs.oracle.com/cd/E11882_01/server.112/e11050.pdf
7. https://www.tutorialspoint.com/distributed_dbms/distributed_dbms_database_environments.htm
8. http://www.techrepublic.com/article/distributed-transactions-span-sql-server-and-oracle/
9. Wang, C.-Y., Spooner, D.L.: Access control in a heterogeneous distributed database management system. In: SRDS (1987)
10. https://docs.oracle.com/cd/B28359_01/server.111/b28310/ds_concepts004.htm#ADMIN12128

Variants of Software Defined Network (SDN) Based Load Balancing in Cloud Computing: A Quick Review

Jitendra Bhatia, Ruchi Mehta(✉), and Madhuri Bhavsar

Nirma University, Ahmedabad 382481, Gujarat, India
mehtaruchi1312@yahoo.com

Abstract. Nowadays users of cloud are increasing rapidly hence handling of and allocation of that resources are the main challenge. Load balancing strategy refers to scatter the dynamic workload over the various node to guarantee that no single node is over-burden. There are few limitation of conventional load balancers in terms of flexibility and adaptability. To overcome this pitfalls, the usage of Software Defined Network based approach in load balancing proves to be advantages. Software Defined Networking is a developing innovation which helps to quickly strategies in familiarizing the administrations with the business segment without relying upon the seller based setup of the gadgets. SDN helps to set control decision for algorithm that apply for system which increases the performance of that algorithm, reduces response time, increase scalability, flexibility and results in reduction of the energy consumption of system. In this paper, we examined the feasibility of SDN-based load balancing and discussed variants of the SDN-based load balancing using various controllers.

Keywords: Cloud computing · Software defined network
Load balancing · OpenFlow · Controller

1 Introduction

1.1 Cloud Computing

The quick improvement of the Internet that has encouraged a large number of new innovations including cloud computing. The cloud computing rapidly emerged as a virtualization technology that aims to provide scalable, transparent network to end users [2]. The cloud can give facility to use on-demand computing and storage application to end users. The end user cannot have knowledge of where that service is from and how they are delivered to them. Cloud Computing has mainly three components that are the client computer, data center, and distributed servers. Client computers means devices that users can communicate with other cloud components such as data center and distributed server via the Internet. There are three types of client that are thin, thick and mobile client.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 164–173, 2018.
https://doi.org/10.1007/978-3-319-73712-6_17

A thin client is a most popular client in the cloud whereas distributed servers are placed at different location. In that central server are there which can monitor traffic, client demands and ensures that all the process runs smoothly or not and the data center is a collection of servers where various application are deployed which can be accessed via the Internet.

As the usage of cloud is increasing, it has a huge effect on the cloud data center because of the large number of request that arrives at the same time, but sometimes the data center cannot allocate the resources at this peak time which results in to the situation for adding the resources to fulfill the end user request. On the off chance that the load is not adjusted at specific path, then including new resources leads to wastage of assets [9]. To accomplish the negligible reaction time and to decrease the utilization of registering assets required that adjusting the heap among all the accessible assets decently.

1.2 Software Defined Network

SDN is a new methodology in networking that allows the administrator of the network to manage the network abstraction through the lower level functionality [22]. The concept of SDN is open and reprogrammable. This is an architecture that is not only controls network devices but also controls an entire network. The main goal of this architecture is to allow network engineers and administrator to respond rapidly to evolving business needs [10]. It control the entire network centrally so that there is no need for touch individual devices for any change in the network. Due to changes of the functions and performance, the traditional networks have lots of drawbacks. For achieving better services we can include so many solutions in the network device. This will make our system large, fat and complex which results in complexity to achieve the better network performance [10]. To overcome this limitation, Software Define network was proposed in the year of 2003 [18]. There are two parts of SDN infrastructure that are control plane and data plane [8]. The control plane is responsible for controlling the data transmitted over the network. The logic of the controlling can be implemented in the server as a software component. And data plane is responsible for forwarding data. It is available on the network devices like switches, router. This SDN is an emerging architecture that is dynamic, cost-effective and manageable. Control plane consists various controller and applications and data plane/forwarding plane consists of various networking devices. OpenFlow protocol is suitable standard protocol for SDN-based system implementation [18].

Software defined network architecture is shown in Fig. 1. A layered architecture that has application layer, the control layer, and infrastructure layer. The application layer consists various applications like firewall, load balancer etc. and network services that interact with the control layer with the help of northbound APIs. The control layer consists of centralizing control plain used for communication with below layer of sdn known as infrastructure layer using OpenFlow protocol which uses southbound API while the infrastructure layer consists both network and virtual devices that implement OpenFlow protocol for implementing the traffic forwarding rule [8].

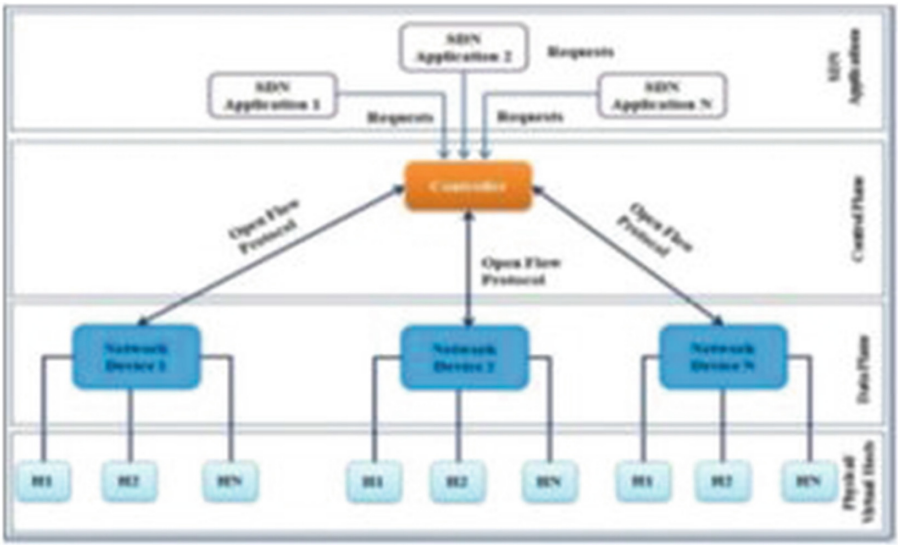


Fig. 1. SDN architecture [8]

1.3 OpenFlow

The OpenFlow standard is utilized as a part of SDN for correspondence between applications. There is likewise an OpenFlow empowered controller characterized in the OpenFlow switch particular. It is supervised by Open Networking Foundation. The OpenFlow convention permits consistency, direct control of the foundation, consequently evacuating the requirement for complex system administration [10]. It includes adaptability and substantial flexibility from the exclusive conventions of a solitary equipment merchant. The design of OpenFlow comprises of three parts: 1. controller, 2. secure channel and 3. OpenFlow empowered switch [10]. Thus the Switches utilize a stream table to forward the bundle to the goal. Stream table contains the rundown of stream sections. Switches utilize the conventions that are characterized in the stream table for sending the parcels. A secure channel is utilized for secure correspondence amongst switch and the controller. The Controller is a product program that is utilized to include, adjust/change and erase the stream table passage of the switch utilizing this convention [10] (Fig. 2).

1.4 Difference Between Traditional Network and SDN Network

The SDN has so many advantages with compared to traditional networking are as followed:

- Migration VM become easier.
- SDN only requires one centralized control plane which offsets the cost of the forwarding plane.

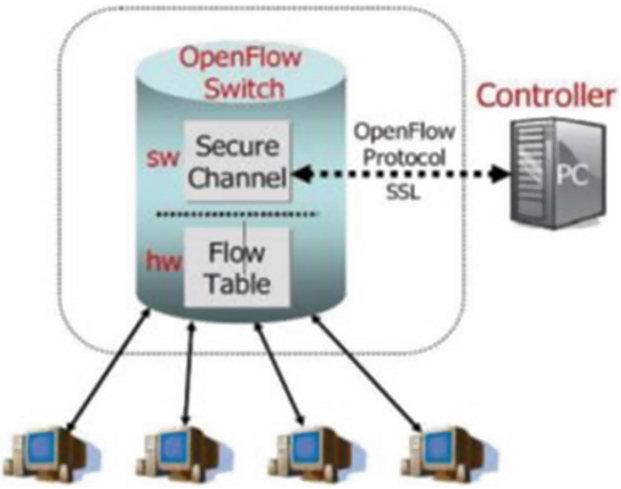


Fig. 2. OpenFlow architecture [8,10]

- In SDN QOS is provided in more efficient way.
- SDN is directly programmable because Whole network is directly programmable because forwarding functions and control functions are logically decoupled from each other which enables the network pragmatically configured by automation tool including OpenStack, Chef, and Puppet.
- SDN does not require a huge amount of resources so that investment of resources can be reduced and also some of the SDN products are open source.

Table 1. Difference between traditional network and SDN network

Traditional networking	SDN
Static and inflexible networks and are not useful for new business ventures	Programmable networks during deployment time as well as at later stage based on the change in the requirements and helps new business ventures through flexibility, agility, and virtualization
Hardware appliances	Configured using open software
Distributed control plane	Logically centralized control plane
Work using protocols	APIs to configure as per need
Policy to treat an incoming data packet is written into its firmware	SDN facilitates admins with granular control over the way switches handle data, giving them the ability to automatically prioritize or block certain types of packets Which in turn, allows for better efficiency without the need to invest in expensive, application-specific network switches

But they also have some challenging issues like Quality of service, security, Load balancing, scalability [8].

1.5 Load Balancing

Load balancing is used to distribute load and improve the performance, availability of resources and scalability of the system, hence we can achieve the minimum response time of the application and increase the throughput [8].

The load balancing algorithm in cloud typically divided into two sections that are static load balancing algorithm and dynamic load balancing algorithm. The static algorithm can most suitable in a homogeneous environment [2]. But this type of algorithm is not suitable where attributes are dynamically changing during the execution time. Whereas dynamic algorithm can be used in the environment when the attributes are changed continuously and also give good result in the heterogeneous and dynamic environment [2].

In the static mechanism, An algorithm like Round robin and central load balancing decision model (CLBDM) [2] are considered. In the Round robin load balancing algorithm, it provides resources to the task as FCFS (First come First serve) basis. The central load balancing decision model works same as round robin load balancing algorithm, but in addition to this, it calculates the duration of the connection of the client request and server response by measuring the overall execution time of the task on the given resources.

In the dynamic load balancing algorithm includes algorithms like Least-connection algorithm, Response time algorithm, and Predictive algorithm [22]. Least-connection algorithm can detect the number of connections that associated between the client and the server, in a particular time interval. And if a new access request arrives then they forward the request to the least connection server [22] and that server processes the request and reply back to the requester. In the Predictive algorithm, the server load can be predicted for the next period of particular requests. In the response time algorithm, the balancer can estimate the load of the each and every server using sending a ping request to the server [22].

These are the traditional network that has some limitations like complex software and hardware used in the system that can increase the operation cost of the organization and their re-usability of code and their applicability to the particular application is different hence architecture is poor.

Table 2. Variants of SDN controllers

Controller	OpenSource	Language	Multi-threaded	GUI	Invented By
Nox [6]	Yes	C++/Python	No	No	Nicira Networks
POX [13]	Yes	Python	–	No	Nicira Networks
Beacon [5]	Yes	Java	Yes	Yes	Standford University
Floodlight [12]	Yes	Java	Yes	Yes	Big Switch Networks
OpenDayLight [7]	Yes	Java	Yes	Yes	Multiple Contributors
RYU [14]	Yes	Python	–	No	NTT OSRG and VA Linux

Mainly in this paper, we surveyed the variants of the current SDN-based load balancer and also discussed the various parameters of that variants and how they can be used to achieve effective load balancing in cloud computing using software define networks.

2 Variants of SDN Based Load Balancer

2.1 Heuristic Based Load Balancer

The main goal of this load balancer is to minimize both server and network load. This method defines its own objective function that discovers best path and best server in the fastest manner. The data plane, controller, application layer is 3 layers of this framework. The data here global view of the perspective network. Hence dynamic path selection can be accomplished by minimal response time than the hop based updates in routing [9].

This method can be tested in the Java environment [9]. Using the objective function complete the request in the smallest span of time. The dynamic path can be reduced, by evaluating hop by hop method of the network. Using objective function, congestion and delay via existing algorithm because can be reduced as it that selects the least loaded path [9].

2.2 SDN Based Traffic Engineering Based Load Balancer

Traffic Engineering (TE) means optimization of performance of the network by dissecting, anticipating and managing the conduct of the information that is transmitted over the network [21]. Software defined network based TE that comprises of ideal design organization and traffic load balancing. Here author describes main component of Traffic Engineering manager for this method [21] that focuses on ideal topology arrangement and traffic load adjusting. The optimal topology consumption is power utilization of DCN is always higher than what amount required because the movement interest in DCN progressively changes from time to time [21]. Traffic load balancing is to minimize the congestion of all the possible connections by separating the traffic of DCN design or subset topology discovery by ideal topology composition algorithm using dynamic traffic load balancing [21]. Here both the algorithm can be describe using heuristic and linear way. This method testing prospective uses the mininet simulator and the virtual instance is OpenVSwitch. For notification Floodlight [12] which is Java API is used. This system reduces the 41% power consumption and 60% lower maximum link utilization compared to existing static routing scheme [21].

2.3 OpenFlow Protocol Based Load Balancer

The main objective of this load balancer changes the manual and costly hardware of clusters to the OpenFlow based controller using the local network infrastructure. The main key idea of this proposed load balancer is to replace the

expensive and statically defined network component and cluster through openflow controller. Here two algorithms such as OpenFlow based Round-Robin and OpenFlow based Least connections algorithm has been proposed [22]. Above two techniques are tested on the mininet [3] and Floodlight [12] Java API. In first approach, response time is very unstable because server cluster varies greatly. Whereas the second approach takes lesser time and give the better performance so that this method always help to forward load to least connection server [22].

2.4 SDN Based Design of Load Balancer Middlebox for Data Center

In this load balancer, there are number of SDN controllers and OpenFlow switches inside the middle box network and that are basically based on clos network [4] framework architecture. Here, Users can be arranged in nonblocking Clos network to achieve better efficiency and resource utilization. There are two methods describes here and that are switches inside the middle box and server inside the middle box [16]. Delay and packet loss can occur because traffic passes inside the middle box through a switch. Port rate, traffic load, queue length etc. are information of the switch that can be collected by a controller. With the help of these attributes the path within the middle box can be changed [16] and load will be equally distributes. This type of SDN load balancer has been tested in the Matlab and it improves the utilization and reduces the latency.

2.5 Data Flow Network Load Balancer in Eucalyptus

Here authors used the Eucalyptus cloud system architecture for proposed LBVMD [19] i.e. load balancer VM deployment mechanism. This system basically consists three components that are Eucalyptus, agent, and the openflow-enabled switch. In that agent is important part of this system because it monitors the network and send information to Cloud controller of Eucalyptus for selecting appropriate Node controller to create new VM [19]. OpenDaylight [7] is used as a controller for the system [19]. In this system, the VM made by the system obliged planning component gives preferred execution over existing instruments. Openflow and the Eucalyptus distributed computing setup are utilized as a part of the testbed [19].

2.6 SDN Based Dynamic Load Balancer

Dynamic load balancing method of cloud-center is based on SDN (SDN-LB). SDN-LB includes four main modules: traffic detection module which is responsible for dynamic traffic monitoring and statistics; load calculation module which aims to estimate the load distribution of cloud environment; dynamic load scheduling module which proposes a hybrid load balance algorithm to realize high performance load balance for Cloud center; flow management module which is responsible for deployment load balance strategy based on a hybrid load balance algorithm [20]. This SDN-based dynamic load balancing algorithm inside

POX [13] controller is used which is written in Python language [20]. It yields a higher throughput and better efficiency compared to existing dynamic load balancing methods [20].

2.7 Extended Health Monitoring for Openflow Network (EHLBOF)

In this paper [17], mainly focuses for checking the health of the server i.e. the status of physical resources of server and applications that are running on the server. The EHLBOF [17] probes periodically and gets the status of servers that are present in the network by using Simple Network Management Protocol (SNMP) request and process the response message from the server and update its status in the server. And suppose any problem arises then discard any future request for that particular server [17]. This method has been tested on the mininet [3] and pox [13] and the result shows that the throughput increases double than the round robin method because this method can dynamically find the status of each server and update the same dynamically [17].

2.8 Path Load Balancing

This paper [11] basically describes path load balancing. This method is divided into three parts that are data collection, evaluation model [11] and flow table installation. This method selects the path dynamically based upon the traffic information of the particular node. This traffic information help to detect fault in link or node based on which they select another path [11]. This method has been tested in mininet [3] and POX [13] and the result when compared with shortest path algorithm in terms of reliable, efficient and effective [11] also it give the assurance about the quality of packet.

2.9 Single Flow Table and Group Flow Table Combination

This paper basically focuses on flow table rules and the algorithms related to single flow table and group flow table and combination of both [15]. Group flow table has the traffic and number of packets. For single flow table is search based on the information of the health of the backend server and in case of rule is utilized directly else it decides whether that is need for modification the load balancing that is matched it offer to monitor in case of server down, maintenance of Group flow table [15]. This method was implemented on the mininet [3] and OpenDayLight [7] and analyzing the results show that the life cycle of single flow table is short and group flow table is very long. Single flow table analyzes traffic of each and every client given its information for changing in the Group Flow table [15].

2.10 End Host Load Balancing

This paper describes [1] mainly two load balancing scheme based on the controller and switch [1]. Controller based load balancer is based upon the Round

Robin in that decision is based on individual TCP session. Here controller can select an interface of the available N interface. In switch-based load balancing, a hash function is used. The Hash function selects the random switch and perform equal load balancing and thereby selecting one interface for whole TCP session [1]. This method has been tested on the GNS3 and RYU [14]. The result was analyzed that controller based load balancing can run in the time that nearest to an optimal value of running time [1].

3 Conclusion

Cloud computing is new emerging era of computing, it uses the resources like processing, storage and network based application. For this type of application, performing the task like customizing systems administration and virtualization are the major issues which can be solved by using SDN. SDN utilizes the system assets flexibly and fulfills the client application without any limitations. We surveyed the SDN based load balancing mechanism in the cloud environment and also discussed various types of variants that helps the load balancing mechanism efficiently with compare to the traditional mechanisms. We concluded from variants that we can achieve minimum response time, higher throughput than conventional method of load balancing and also results in reduction of the power consumption using traffic engineering mechanism. Using OpenFlow, it forwards the load to least connection sever which reduces latency and improves the utilization of data center using close network based middle box design. It improves server health problem and also finds an optimal value of running time using SDN controller based load balancer.

References

1. Al-Najjar, A., Layeghy, S., Portmann, M.: Pushing SDN to the end-host, network load balancing using openflow, pp. 1–6 (2016)
2. Al Nuaimi, K., Mohamed, N., Al Nuaimi, M., Al-Jaroodi, J.: A survey of load balancing in cloud computing: challenges and algorithms, pp. 137–142 (2012)
3. Antonenko, V., Smelyanskiy, R.: Global network modelling based on mininet approach, pp. 145–146 (2013)
4. Clos, C.: A study of non-blocking switching networks. Bell Syst. Tech. J. **32**(2), 406–424 (1953)
5. Erickson, D.: The beacon openflow controller, pp. 13–18 (2013)
6. Gude, N., Koponen, T., Pettit, J., Pfaff, B., Casado, M., McKeown, N., Shenker, S.: Nox: towards an operating system for networks. ACM SIGCOMM Comput. Commun. Rev. **38**(3), 105–110 (2008)
7. https://www.opendaylight.org/. The opendaylight platform
8. Meng, K.C., Govindarajan, K., Ong, H.: A literature review on software-defined networking (SDN) research topics, challenges and solutions. In: 2013 Fifth International Conference on Advanced Computing (ICoAC) (2013)
9. Koushika, A., Selvi, S.: Load valancing using software defined networking in cloud environment. In: 2014 International Conference on Recent Trends in Information Technology (2014)

10. Lara, A., Kolasani, A., Ramamurthy, B.: Network innovation using openflow: a survey. IEEE Commun. Surv. Tutor. **16**(1), 493–512 (2014)
11. Li, J., Chang, X., Ren, Y., Zhang, Z., Wang, G.: An effective path load balancing mechanism based on SDN, pp. 527–533 (2014)
12. Big Switch Network. http://www.projectfloodlight.org
13. Nicra. http://www.noxrepo.org/pox/about-pox/
14. NTT Laboratories OSRG Group NTT. http://osrg.github.com/ryu/
15. Qilin, M., Shen, W.: A load balancing method based on SDN, pp. 18–21 (2015)
16. Zhao, J., Yang, Y., Shi, L., Tu, R., Wang, X., Wolf, T.: Design of a load-balancing middlebox based on SDN for data centers. In:2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (2015)
17. Saifullah, M.A., Maluk Mohamed, M.A.: Open flow-based server load balancing using improved server health reports, pp. 649–651 (2016)
18. Network Virtualization. https://www.sdxcentral.com/sdn/definitions/what-the-definition-of-software-defined-networking-sdn/
19. Chen, J., Chou, F., Hsieh, W., Hsieh, W., Lee, Y.: Load balancing virtual machines deployment mechanism in sdn open cloud platform. In: 2015 17th International Conference on Advanced Communication Technology (ICACT) (2015)
20. Qian, H., Yong, W., Xiaoling, T., Yuwen, K.: A dynamic load balancing method of cloud-center based on SDN. China Commun. **13**, 130–137 (2016)
21. Li, J., Hyun, J., Yoo, J., Han, Y., Seo, S., Hong, J.: Software defined networking-based traffic engineering for data center networks. In: The 16th Asia-Pacific Network Operations and Management Symposium (2014)
22. Zhang, H., Guo, X.: SDN-based load balancing strategy for server cluster. In: 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems (2014)

Analysis of Ionospheric Correction Approach for IRNSS/NavIC System Based on IoT Platform

Mehul V. Desai(✉) and Shweta N. Shah

Electronics Engineering Department, SVNIT, Surat 395007, Gujarat, India
mvd.svnit@gmail.com, snshah@eced.svnit.ac.in

Abstract. The supreme success of the future Internet of Things (IoT) depends on the ubiquitous and immaculate connectivity provides by satellite. Ionosphere is one of the major contributing factor in signal propagation for satellite based application, which results in degradation of measurement accuracy. In the India, Indian Regional Navigation Satellite System (IRNSS)/Navigation with Indian Constellation (NavIC) both *L5* and *S* band signals are more affected by this ionosphere due to its low latitude geographical location. So, the future success of IRNSS system based on IoT platform depends on accuracy of ionospheric mitigation algorithm. This paper concentrate on comparative analysis of coefficient based model and dual frequency model based ionospheric model. The data is collected from IRNSS/NavIC receiver located at communication research laboratory, Electronics Engineering Department, SVNIT surat (21.16° Lat, 72.78° Long) provided by SAC, ISRO Ahmedabad. It is observed that the amount of delay contribution by *L5* band is more compared to *S* band. The performance of dual frequency and coefficient based model is checked on different geomagnetic *Kp* index. It is also deduced from the comparison that the dual frequency model works better in stormy days, where coefficient based approach gave bad performance.

Keywords: Indian Regional Navigation Satellite System (IRNSS)
Navigation indian constellation (NavIC) · Ionodelay
Grid iono vertical error (GIVE) · Klobuchar · Dual frequency

1 Introduction

The goal of the IoT is that all devices should be connected wherever they are located. Where Wi-Fi, Bluetooth and GSM networks are fail to provides the ubiquitous and seamless coverage services there satellites works better. Hence, The ultimately future of IoT will depend on the satellite based network [1]. The satellites network provides a good Coverage, high reliability, low latency, high speed, versatile and cost effective services [2]. Integrating IoT with satellite system will solved many problem of navigation e.g. transportation problem like, traffic jam, road block etc. The success of satellite based navigation application depends on accuracy of measurement and it is noticed that measurements always affected by different error sources.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 174–183, 2018.
https://doi.org/10.1007/978-3-319-73712-6_18

Indian Space Research Organization (ISRO) developed IRNSS/NavIC, which will gives positioning service with a 10 m of accuracy for both civilian and military users of the India [3]. The IRNSS consists three Geostationary Earth Orbit (GEO) and four Geo Synchronous Orbit (GSO) satellites [4]. The arrangement of three GEO is done at 32.5 °E, 83 °E and 131.5 °E longitude and four GSOs are in two planes that cross the equator at 55° and 111.75° East respectively. The IRNSS satellites broadcast the signal in L_5 band (1164.45 1188.45 MHz) and S band (2483.5–2500 MHz) with a carrier frequency of 1176.45 MHz and 2492.08 MHz respectively [4, 5]. The military or defense signal is encoded and modulated by Binary Offset Carrier (BOC) (5,2) for secure communication. In contrast with it, the civilian signal is simply used Binary Phase Shift Keying (BPSK) modulation [4, 6]. Currently, the IRNSS fully operational as all seven satellites, 1A, 1B, 1C, 1D, 1E, 1F and 1G are available in an orbit [7]. IRNSS/NavIC satellites consist navigation and ranging payload. The IRNSS/NavIC users compute their position by the navigation signal provides by the receiver.

The IRNSS/NavIC both *L5* and *S* band signals passes through the atmosphere before reaching the user receiver, thus the signals are always affected by different error sources whether it is intentional or unintentional [8]. Hence the position computed by IRNSS/NavIC users is always deviated. The ionosphere with an altitude between 60 km to 700 km above the earth's surface contribute highest error in position measurement by IRNSS/NavIC. The behavior of ionosphere is irregular when the earth's magnetic field is disturbed, geomagnetic storm and mass ejection of the solar corona is occurred [9]. In India as the large irregularities are available in ionosphere IRNSS/Navic both signals are highly affected by it. To mitigate this error due to ionospheric irregularities, different methods is applied like, dual frequency methods, differential correction approach and various single frequency ionodelay models. In this paper performance investigation of eight coefficients (four α and four β) based model [10, 11], dual frequency model is done for ionospheric correction on IRNSS/Navic receiver. In IRNSS/NavIC users can apply the ionospheric correction by three ways (i) grid based (ii) coefficient (iii) dual frequency. The IRNSS/NavIC is broadcasting, 8 correction coefficient of coefficient based model and 80 Ionospheric Grid Point (IGP)correction for GIVE model in *L5* band signal [4]. Detail information related to coefficient based and dual frequency model is described in the Sect. 1. Section 2 contains the analysis of all ionospheric model. Finally, conclusion of this paper is included.

2 Ionospheric Correction Models

The amount of delay contributed by the ionosphere depends on density of free electron present on it, called Total Electron Content (TEC). The TEC density is changed during day and night time due to recombination and ionization process. It is also depends on seasonal behavior condition and solar cycle and geographical location of the user [12]. The quiet and stormy days are identified by a variety of geomagnetic indices, such as K, K_p, A_p and D_{st}, and it is correlated with the variation of TEC in the ionosphere [13]. There is a large gradient observed in ionosphere near Indian region. Hence, the IRNSS performance only succeeds when these effects will be mitigated effectively using some models or method in real time scenario. In a matured GPS system normally coefficient

based (klobuchar model) is applied for the correction [14]. Here coefficient based correction is also applied on both the bands of IRNSS, which is explained below.

Ionodelay Computation Using Coefficient. The master frame of IRNSS contains four sub frames and each sub frame is 600 symbols long, so total 2400 symbols per frames [4, 7]. Sub frames 1 and 2 transmit primary and sub frames 3 and 4 transmit secondary navigation parameters respectively [3]. Secondary navigation parameters include, ionospheric delay correction coefficient and Ionospheric grid delays and confidence values. The Ionodelay computation using coefficient is empirical model and estimated the delays based on 8 coefficient [10, 14], which are broadcasted through navigation data once in a day. The steps for the algorithm is as follows

Algorithms

Step-1: Using Azimuth (A_z) and Elevation (El) angles, compute Earth's central angle (Ψ) in semi-circles [4, 14].

$$\Psi = \frac{0.0137}{El + 0.11} - 0.022 \tag{1}$$

Step-2: Compute geodetic latitude (ϕ_i) and longitude (λ_i) of the earth projection intersection point of ionosphere in semi-circles [10].

$$\phi_i = \phi_u + \psi \cos A_z \qquad |\phi_i| \leq 0.416 \tag{2}$$

if $\phi_i > +0.416$ then $\phi_i = +0.416$
if $\phi_i < -0.416$ then $\phi_i = -0.416$

$$\lambda_i = \lambda_u + \frac{\Psi \sin A_Z}{\cos \phi_i}$$

where, λ_u and ϕ_u are user's geodetic longitude and latitude in semi-circles respectively.

Step-3: By assuming mean ionospheric height h 350 km geomagnetic latitude at point where projection of earth intersect with ionosphere is calculated by [4, 11, 14]

$$\phi_m = \phi_i + 0.064 \cos(\lambda_i - 1,617) \tag{4}$$

Step-4: After correction coefficient (α,β) received from satellites, compute the amplitude and delay of the ionospheric delay denoted as A_I and T_I [4].

$$A_I = \sum_{n=0}^{3} \alpha_n \phi_m^n \quad A_I \geq 0 \tag{5}$$

if $A_I < 0$ then $A_I = 0$

$$T_I = \sum_{n=0}^{3} \beta_n \phi_m^n \quad T_I \geq 72,000 \tag{6}$$

if $T_I < 72,\ 000$ then $T_I = 72,\ 000$ (sec) and depending on T_I value parameter x is derived by

$$x = \frac{2\pi(t - 50,400)}{T_I}$$

Where, t is calculating as,

$$t = \left(4.32 * 10^4\right) * \lambda_i + TOWC(IRNSS)$$

Step-5: Depending on this x parameter value ionospheric correction is applied as [4, 14].

If, $|x| < 1.57$ then

$$T_{iono} = F * \left[5.0 * 10^{-9} + AMP\left(1 - \frac{x^2}{2!} + \frac{x^2}{4!}\right)\right] \tag{7}$$

otherwise

$$T_{iono} = F * \left(5.0 * 10^{-9}\right) \tag{8}$$

Coefficient model is very simple, As the coefficients are fixed for a day, it can not work efficiently. Compare to that dual frequency model is more efficient which is explained next.

2.1 Dual Frequency Model

Instead of using coefficient based single frequency ionospheric correction model for estimation of ionodelay at user's location, another method can be adopted. This method uses NavIC/IRNSS pseudo-range measurement at both L_5 and S frequencies. The TEC is computed and converted into ionodelay in meter using conversion factor. The two frequencies L_5 and S user shall correct for the group delay due to first order ionospheric effects by applying the relationship [4]:

$$\sigma = \frac{\sigma_{L_5} - \gamma * \sigma_s}{1 - \gamma}, \tag{9}$$

where, denoting the nominal center frequencies of L_5 and S respectively,

$$\gamma = \left(f_s^2 / f_{L_5}^2\right), \tag{10}$$

where, σ = pseudorange corrected for first order ionospheric effects. σ_{L5}, σ_S = pseudorange measured on the channel indicated by the subscript. The comparative analysis between dual frequency and single frequency model is included in below section.

3 Simulation and Results Discussion

Ionospheric delay estimation for NavIC/IRNSS is carried out based on MATLAB. The simulation flow diagram is depicts in Fig. 1. The one week data starting, which is start on Sunday and end at Saturday have been used for analysis. The one week raw data of IRNSS/NavIC satellites starting from Time Of Week Count (TOWC) 0 (starting of the Sunday) to 648000 (end of the Saturday) is collected by the IRNSS/NavIC receiver at communication research laboratory, SVNIT, Surat (21.16° Lat, 72.78° Long). Ranges between IRNSS satellites (1A, 1B, 1C, 1D, 1E, 1F, 1G) and user receiver is calculated by extracting primary information from the raw data. First ranges for both L_5 and S band are calculated then dual frequency approach [8] is applied to measure the ionodelay for individual satellite.

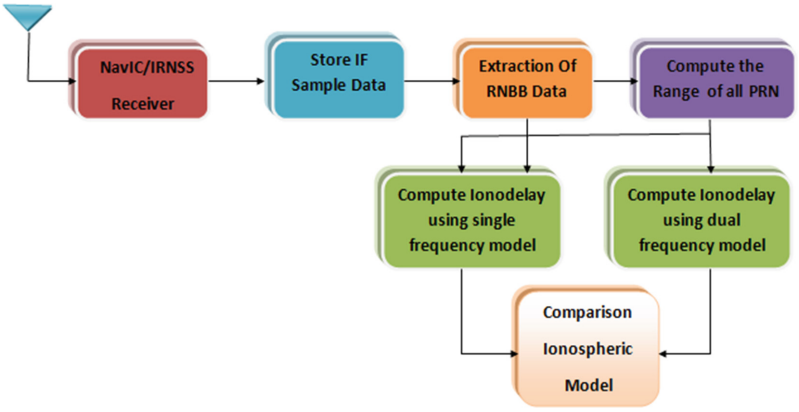


Fig. 1. Block diagram of simulation setup

Figure 2 shows the comparisons of ionodelay calculated by dual frequency approach for IRNSS six satellites namely 1B, 1C, 1D, 1E, 1F and 1G on the stormy day 14/08/16 ($3 < K_P < 5$). The observation was carried out for individual six satellites and it is observed that all individual satellites have a large ionodelay in *L5* band compared to *S* band. As per the literature maximum ionodelay will happen when the ionosphere recombination rate is lowest. And for the low latitude Indian region, it will happen in

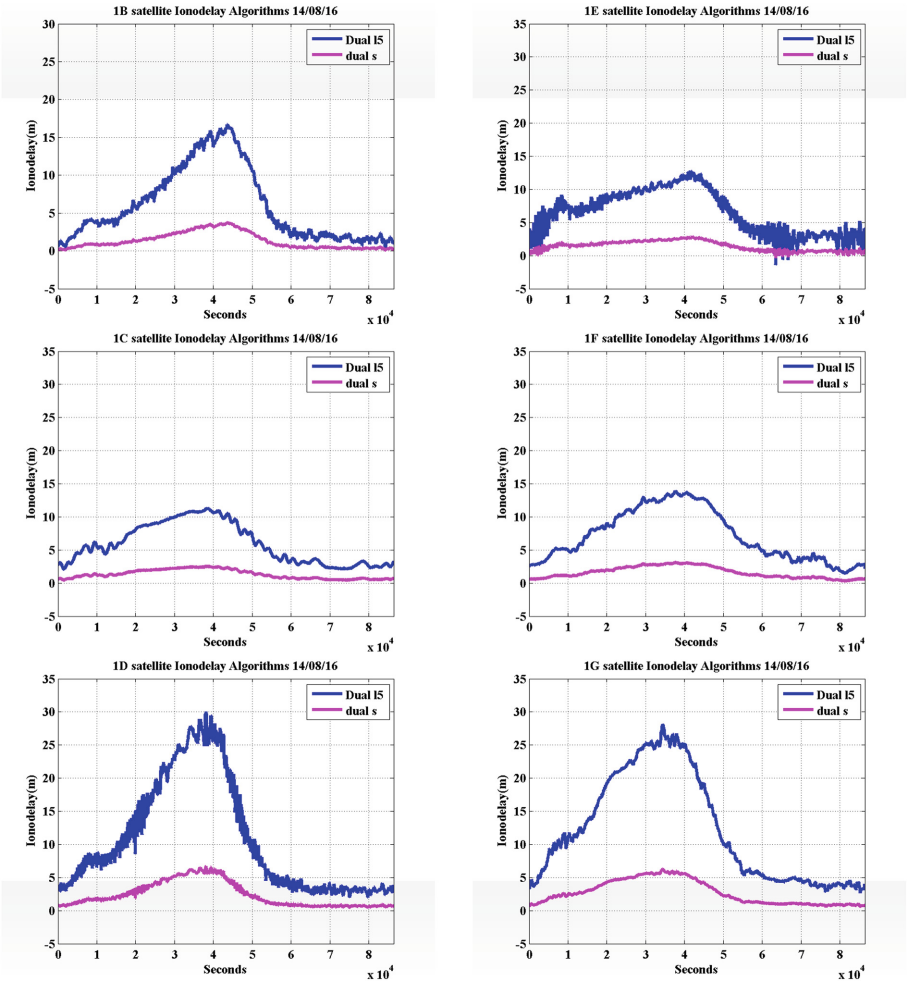


Fig. 2. Ionodelay computed by the dual frequency model on a quiet day 14/08/16 ($3 < K_P < 5$)

the afternoon around period (12.00 to 14.00 h). It is also found from the comparison that maximum ionodelay is estimated at Local Time (LT) around 14.00 (hours).

Ionodelay shown in Fig. 1 is compared in term of maximum, mean and stand deviation values, which are listed in Table 1. The value of maximum ionodelay in meter are 18.0632 m, 11.9698 m, 30.8831 m, 13.4083 m, 14.7313 m, and 29.3238 m for satellites 1B, 1C, 1D, 1E, 1F and 1G respectively. It is noticed that Maximum ionospheric effect felt by satellites 1B and 1D satellites. 1D satellite has a higher delay among all the satellites and its value for L_5 and S bands are **30.8831** m and **06.8828** m respectively. In the case of mean value 1G satellite have a highest value 12.1734 m for L_5 and 02.7130 m for S band. Similarly for the comparison of standard deviation 1G satellites in L_5 (08.1183 m) and 1D satellite in a S(01.8767 m) band have a higher value.

Table 1. Detail ionospheric delay comparisons computed by dual frequency model on 14/08/16 ($3 < K_P < 5$)

Satellites	1B	1C	1D	1E	1F	1G
L5 band dual frequency approach (14/08/16)						
Maximum(m)	18.0632	11.9698	**30.8831**	13.4083	14.7313	**29.3238**
Mean (m)	6.0113	5.9133	**10.6249**	6.4330	7.1613	**12.1734**
Standard deviation (m)	4.8206	3.0610	**8.4209**	3.3964	3.8722	**8.1183**
S Band Dual Frequency Approach (14/08/16)						
Maximum(m)	4.0257	2.6676	**6.8828**	2.9882	03.2832	**06.5352**
Mean (m)	1.3397	1.3179	**2.3679**	1.4337	1.5960	**2.7130**
Standard deviation (m)	1.0743	0.6822	**1.8767**	0.7569	0.8630	**1.8093**

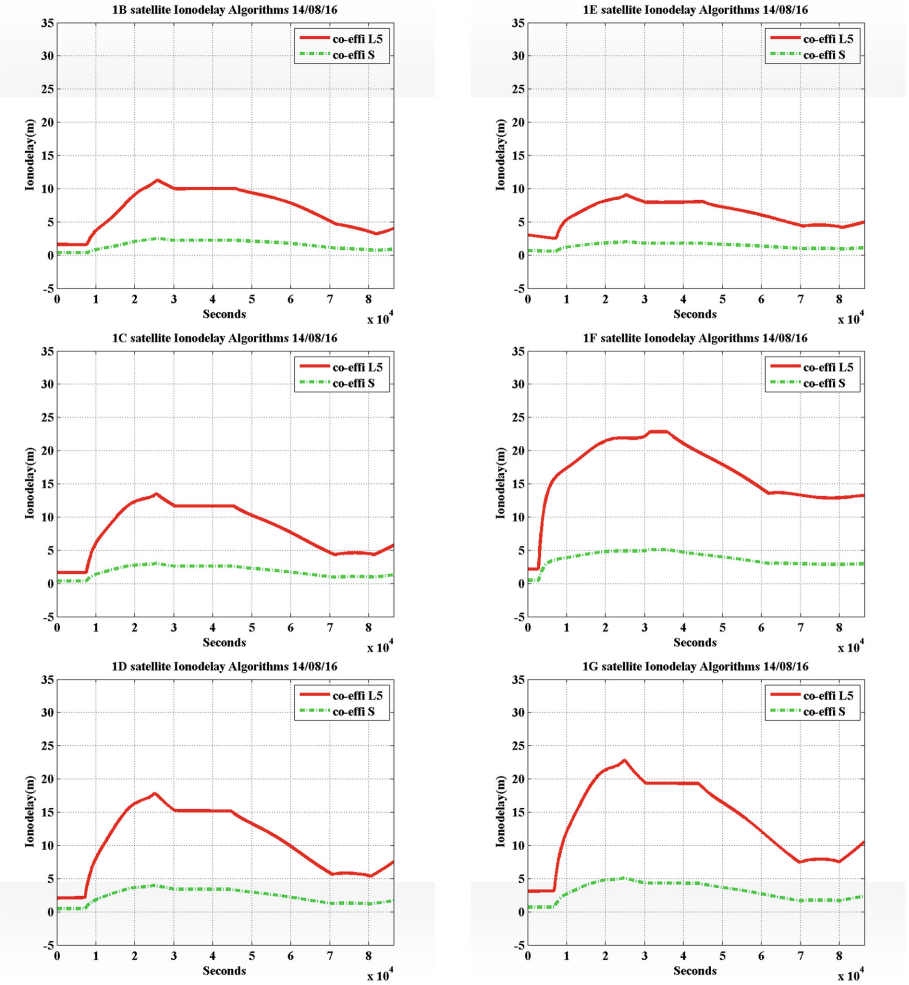


Fig. 3. Ionodelay computed by the 8 coefficient based model on a quiet day 14/08/16 ($3 < K_P < 5$)

As mentioned in literature dual frequency approach gives always better performance, but it has a cost of additional frequency. Hence, the single frequency ionodelay model is applied for comparison. To apply coefficient based model, The broadcasted ionospheric correction coefficients are extracted from raw data. The detail performance comparison of coefficient based model for both band are done for 14/08/16, which is shown in Fig. 3. It has been observed that in coefficient based model cases also L_5 band signal suffers more delay compared to S band signal. The detail comparison is listed in Table 2.

Table 2. Detail ionospheric delay comparisons computed by 8 coefficient model on 14/08/16 ($3 < KP < 5$)

Satellites	1B	1C	1D	1E	1F	1G
L5 band eight coefficient based model (14/08/16)						
Maximum(m)	11.3029	13.4974	17.8568	9.0782	**22.8248**	**22.8896**
Mean (m)	6.9525	8.1487	10.6045	6.2006	**16.7774**	**13.7900**
Standard deviation (m)	3.0082	3.5927	**4.7369**	1.8211	4.6225	**5.8495**
S band eight coefficient based model (14/08/16)						
Maximum(m)	2.5190	3.0081	3.9796	2.0233	**5.0868**	**5.1013**
Mean (m)	1.5495	1.8160	2.3634	1.3819	**3.7391**	**3.0733**
Standard deviation (m)	0.6704	0.8007	**1.0557**	0.4059	1.0302	**1.3037**

It has been observed from the comparison that dual frequency approach has a maximum delay for 1D satellite and its value is around 30.8831, while the coefficient based model have the value 17.8568. So, the coefficient based model perform worst compared to dual frequency model. The performance of the dual and coefficient model also checks for another stormy day ($K_P > 5$) 16/08/2016 where large iono-gradient present. This comparison is shown in Fig. 4.

It has been found that coefficient based model correct only around 50% ionospheric delay correction compared to the dual frequency. The detail comparison is covered in Table 3. Here also noticed that the coefficient based model has failed to compute better

Table 3. Detail ionospheric delay comparisons computed by dual frequency model and 8 coefficient model on 16/08/16 ($5 < KP < 7$)

Satellites	1B	1C	1D	1E	1F	1G
L5 band dual frequency approach (16/08/16)						
Maximum(m)	22.3615	16.8779	**28.0752**	21.5387	21.1992	**29.9452**
Mean (m)	8.0338	8.2158	**11.6218**	9.8453	9.9320	**15.2298**
Standard deviation (m)	6.6234	5.2157	**7.9698**	6.4735	6.3990	**9.3670**
L5 band eight coefficient based model (16/08/16)						
Maximum(m)	10.0792	11.6613	15.2376	8.1150	**22.8218**	**19.3657**
Mean (m)	7.5289	8.6454	11.2386	6.4872	**16.6324**	**14.4466**
Standard deviation (m)	2.3803	2.7243	**3.6256**	1.3776	3.3636	**4.4180**

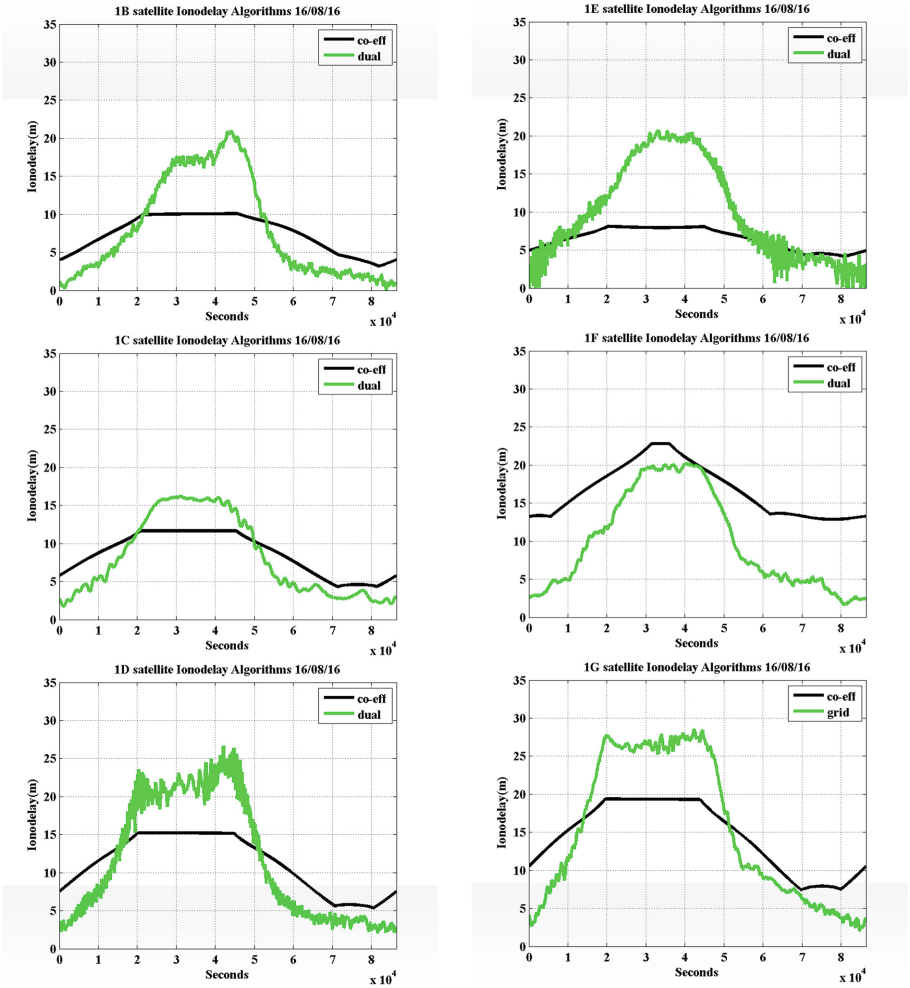


Fig. 4. Ionodelay computed by the dual frequency and 8 coefficient based model on a day 16/08/16 ($K_P > 5$)

ionodelay in both stormy days. Where, the dual frequency model perform good in stormy days also.

4 Conclusion

The paper contains a comparative analysis of different ionospheric models for future NavIC/IRNSS system based IoT platform. The comparison is done between dual frequency method with single frequency eight coefficient model. It has been observed from the dual frequency analysis that *L5* band signal gets more affected by ionosphere

compared to *S* band signal. The maximum error contributed by ionosphere for *L5* band signal is around 30 m at LT 14.00 h. To reduced the cost of extra frequency conventional eight coefficient single frequency model is applied. However, the coefficient base model provides around only 57% correction for a quiet day 14/08/16 and for a stormy day 16/08/16 it's performance is worst. It has been deduced from the comparison that in the both cases dual frequency models gives good performance but with the cost of extra frequency.

Acknowledgments. We are very thankful to Director, Group head DCTG SNAA and Scientist of Space Application Center, ISRO, Ahmedabad for providing the necessary guidance to do this type of analysis on IRNSS receiver provided by them.

References

1. Zhan, H., Wen, Z., Wu, Y., Zou, J., Li, S.: A GPS navigation system based on the internet of things platform: In: IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS), pp. 160–162 (2011)
2. Can the Internet of Things (IoT) Survive without Satellite? Thuraya. http://www.thuraya.com
3. Thoelert, S., Montenbruck, O., Meurer, M.: IRNSS-1A: signal and clock characterization of the Indian Regional Navigation System. GPS Solut. **18**(1), 147–152 (2014)
4. IRNSS SIS ICD for SPS: ISRO-ISAC V 1.1, April 2011
5. Desai, M.V., Jagiwala, D., Shah, S.N.: Impact of dilution of precision for position computation in indian regional navigation satellite system. In: IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, pp. 980–986 (2016)
6. Ruparelia, S., Lineswala, P., Jagiwala, D., Desai, M.V., Shah, S.N., Dalal, U.D.: Study of L5 Band Interferences on IRNSS: International GNSS (GAGAN-IRNSS) User Meet, Bengaluru (2015)
7. Indian Space Research Organization, Applications, Satellite-Navigation Program. http://www.isro.gov.in
8. Desai, M.V., Shah, S.N.: Ionodelay models for satellite based navigation system. Afr. J. Comput. ICT **8**(2), 25–32 (2015)
9. Space Weather Prediction Center, National Oceanic and Atmospheric Administration. http://www.swpc.noaa.gov/phenomena/coronal-mass-ejections
10. Rethika, T., Nirmala, S., Rathnakara, S.C., Ganeshan, A.S.: Ionospheric delay estimation during ionospheric depletion events for single frequency users of IRNSS. Innov. Syst. Des. Eng. **6**(2), 98–107 (2015)
11. Rethika, T., Mishra, S., Nirmala, S., Rathnakara, S.C., Ganeshan, A.S.: Single frequency ionospheric error correction using coefficients generated from regional ionospheric data for IRNSS. Indian J. Radio Space Phys. **42**, 125–130 (2013)
12. Panda, S.K., Gedam, S.S., Jin, S.: Ionospheric TEC variations at low Latitude Indian Region. INTECH Open science (2015)
13. Venkata Ratnam, D., Sarma, A.D., Satya Srinivas, V., Sreelatha, P.: Performance evaluation of selected ionospheric delay models during geomagnetic storm conditions in low latitude region. Radio Sci. **46**(3) (2011)
14. Klobuchar, J.A.: Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron. Syst. **3**, 325–331 (1987)

FFT Averaging Ratio Algorithm for IRNSS

Sreejith Raveendran(✉), Mehul V. Desai, and Shweta N. Shah

Electronics Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat 396007, Gujarat, India
sreejith236@gmail.com, mvd.svnit@gmail.com, snshah@eced.svnit.ac.in

Abstract. Navigation with Indian Constellation (NavIC) or Indian Regional Navigational Satellite System (IRNSS) is an independent satellite based navigation system developed by Indian Space Research Organization (ISRO). Due to solar activity the Total Electron Content (TEC) of atmosphere will have fluctuations, which causes fluctuations in the satellite signal. IRNSS based on Internet of Things (IoT) platform positioning system signals also experience delays as it propagates through the atmosphere irregularities, the majority of which is contributed by the ionosphere. Hence, to make IRNSS based application free of ionodelay prior detection technique is essential. In this paper analysis is done using Fast Fourier Transform (FFT) Averaging Ratio (FAR) to classify the satellites which are more affected by the ionospheric irregularities. The detection threshold is found by using an inverse chi-squared distribution. The data received at the IRNSS receiver located at Communication Research Lab, SVNIT Surat (21.16^0 N, 72.78^0 E), Gujarat is considered for analysis.

Keywords: Indian Regional Navigation Satellite System (IRNSS)
Ionodelay · Total Electron Content (TEC)
Fast Fourier Transform (FFT) · FFT Averaging Ratio (FAR)

1 Introduction

IRNSS is a satellite based regional navigational system consisting of 3 geo-synchronous and 4 geo-stationary satellites developed by ISRO, India. The applications of such a system includes, but is not limited to: Intelligent navigation system based on Internet of Things (IoT). The accuracy of such a positioning system is depending on many factors. Since the positional accuracy of the system depends on the time delay of arrival of the signal from the satellite, even a small delay in the signal offsets the position of user (receiver) by a significant amount. There are different sources of error in pseudorange measurements [1]. The major source of error in pseudorange measurement is caused by the ionosphere [2].

The ionosphere, due to sun's energy, is having a large portion of free electron and charged ions. Since the radio waves are electromagnetic in nature, they are affected by the charged ionosphere in varying manner. Since the ionosphere is

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 184–191, 2018.
https://doi.org/10.1007/978-3-319-73712-6_19

a dispersive medium (delay is different for different frequency), the ionospheric delay can be estimated proportional to the TEC by using [3,4].

$$\text{TEC} = \frac{1}{40.3}\frac{F_1^2.F_2^2}{(F_1^2 - F_2^2)}(\text{P1} - \text{P2}) \tag{1}$$

where P1 and P2 are pseudoranges measured in two frequencies F1 and F2.

The ionospheric delay is directly proportional to the TEC over the region through which the signal propagates. The ionospheric irregularities are more prominent in the latitude range of $\pm 15^0$ to $\pm 20^0$ [5]. Indian territory in this area experiences TEC variations at a higher scale. Due to the relative location of the IRNSS satellite, signal from each satellite experiences a different amount of ionospheric delay. The solutions obtained by considering those satellites with higher ionospheric irregularities tend to be inaccurate. By avoiding these satellites, positioning accuracy can be improved.

In this paper classification of satellite, which are more affected, is done by using FAR Algorithm. The classified satellites can be removed or ignored during computation of Position Velocity Time (PVT) measurements, if sufficient number of satellites are available. The algorithm calculates a decision variable for each satellite and then is classified on the basis of a threshold, set using an inverse chi-squared distribution function [6]. The algorithm is usually used to detect traffic in radio channels used by Chen et al. [7], which is explained in next section, followed by the results and observations with the test data.

2 FAR Algorithm

IRNSS satellites that were affected by the ionospheric effects were evaluated using the FAR algorithm. For the analysis the TEC data from each satellite was taken and given as input to the algorithm. Let $a_n(t)$ be the variation in TEC, where the data is taken at 1 sample per second. Applying FFT to the input, the output is obtained as [6]

$$A_n(s) = \sum_{l=0}^{L-1} a_k\,(l)\, e^{-j2\pi sl/L} \tag{2}$$

where l = 0 to L−1 and n = 1 to N, the number of satellite.

Consider that the input data sequence is even numbered samples. Since the FFT output is symmetric, only the first half of the output $A_n(s)$ is considered. So s = 0 to L/2 +1 and n = 0 to N.

Next the Power Spectral Density (PSD) is found by squaring each of the terms.

$$PSD_n\,(s) = |A_n\,(s)|^2. \tag{3}$$

For every segment, the average of all values of PSD is calculated by using

$$PSD_{avg}(n) = \frac{1}{L}\sum_{s=1}^{L} PSD_n(s). \tag{4}$$

The mean of PSD corresponding to each PRN is calculated by using

$$PSD_{mean} = \frac{1}{N}\sum_{n=1}^{N} PSD_{avg}(n). \tag{5}$$

Now a decision criteria is defined by taking the ratio between P_{avg} and P_{mean}

$$D(n) = \frac{PSD_{avg}(n)}{PSD_{mean}}. \tag{6}$$

In order to classify the satellite, the decision criteria is compared with the threshold. The classification is done by the following rule

$$D(n) \underset{Quiet}{\overset{Disturbed}{\gtrless}} \alpha \tag{7}$$

where α is the threshold value. The threshold is calculated using the inverse chi-squared distribution [8].

The implementation of FAR algorithms on IRNSS system is covered in next section

3 Results and Observations

The simulation related to FAR algorithm on IRNSS data is done on MATLAB 2014 platform. Flow graph for this simulation is summarized in Fig. 1

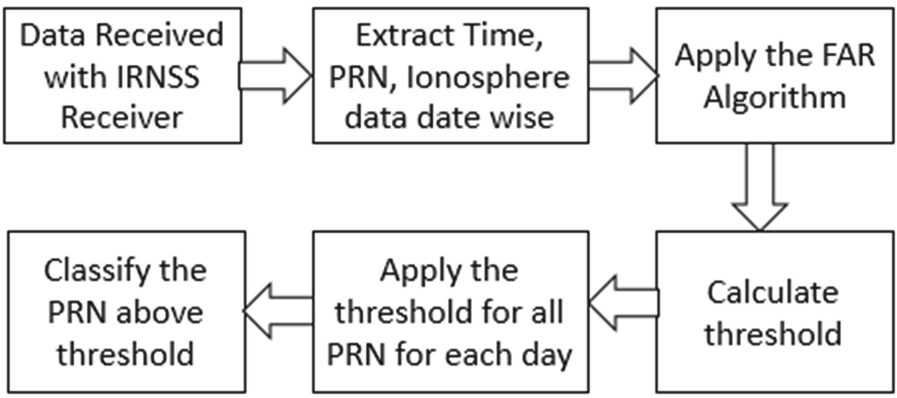


Fig. 1. Flowgraph of processing

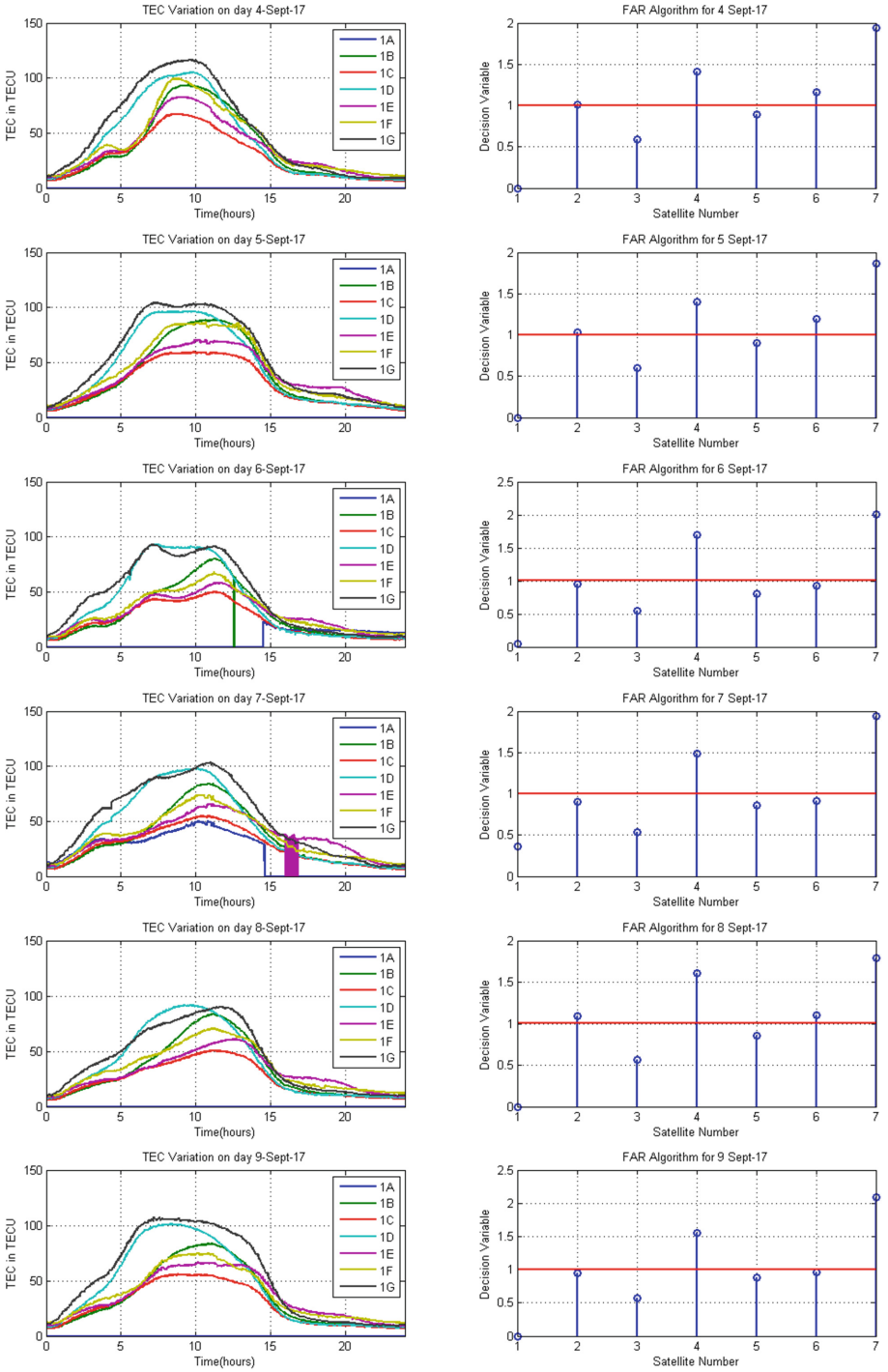


Fig. 2. FAR algorithm for 2nd week of September 2016

The necessary data was obtained from the IRNSS receiver. The data considered was for 2nd week of September 2016. The data was sorted day wise with each of the 7 satellite having their own corresponding TEC data. The TEC data was plotted for 6 days and the FAR algorithm applied on each satellite for 6 days and plotted. The TEC data and the result of the algorithm are plotted side by side as shown in Fig. 2.

In each day, the different satellites- IRNSS 1A, 1B, 1C, 1D, 1E, 1F and 1G provide different TEC values. The shape of the curves are similar- it all reaches a peak from a low value and then gradually decreases to that low value. But the peak of each satellite is different. It is noticed that in all days, the highest peak is that of 7th satellite (1G). The lowest peak is corresponding to 3rd satellite(1C). The data of 1A satellite should be neglected as the receiver was not tracking the satellite properly.

The decision variable obtained after applying the algorithm is in correlation with the TEC data. The highest value of decision variable is for the 1G satellite and the lowest one for 1C satellite.

Table 1. Mean, Minimum, Maximum values of TEC averaged for 6 days

Parameter	1A	1B	1C	1D	1E	1F	1G
Mean	4.4229	34.5293	**27.1646**	42.6009	34.3781	37.1597	**49.5688**
Minimum	0	0	6.1590	6.8767	0	9.0821	8.3974
Maximum	49.7571	93.6586	**67.1377**	105.1793	82.7566	99.1519	**116.4038**

From Table 1 it is observed that 1G satellite is experiencing much higher ionospheric effects compared with the others. The data is obtained by taking mean over the entire day and the averaging over the 6 observed days. The min value of some satellites are shown as zero. This is because of receiver losing track of the respective satellite. Neglecting the 1A satellite, the 1C satellite has the

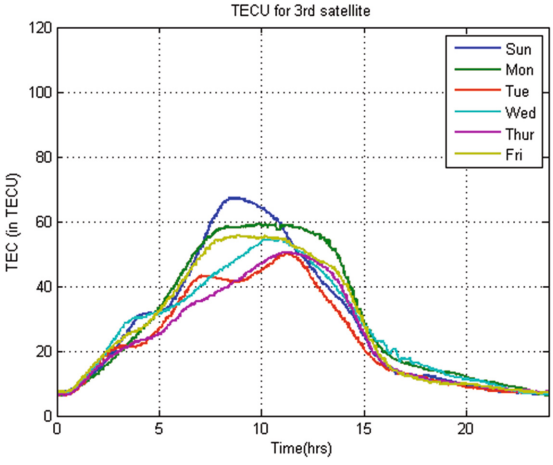


Fig. 3. TEC data observation for 3rd satellite on different days

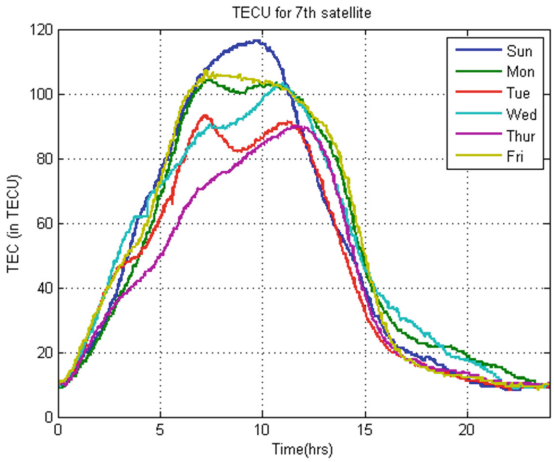


Fig. 4. TEC data observation for 7th satellite on different days

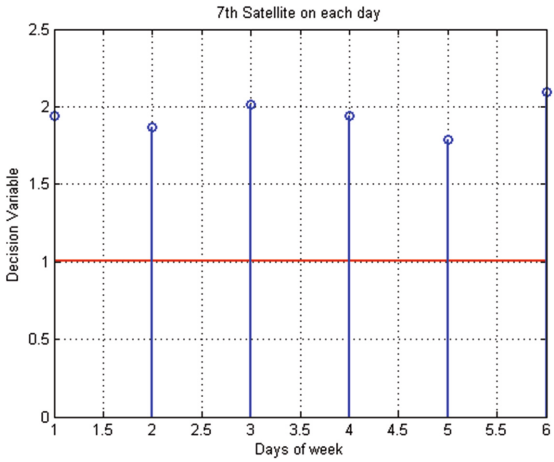


Fig. 5. Decision for 7th satellite

lowest of mean and maximum values for TEC and 1G is having the highest of mean and maximum. The TEC data of the 3rd and 7th satellites are shown in Figs. 3 and 4 respectively for clarity.

Comparing Figs. 3 and 4 it can be seen that the ionospheric irregularities experienced by the 7th satellite is much higher than that of the 3rd.

It can be seen from Fig. 5, the 7th satellite is always classified as experiencing a stormy ionosphere, whereas the 3rd satellite (Fig. 6) is always classified as a quiet ionosphere. This is agreeing with the TEC data obtained for the corresponding satellite.

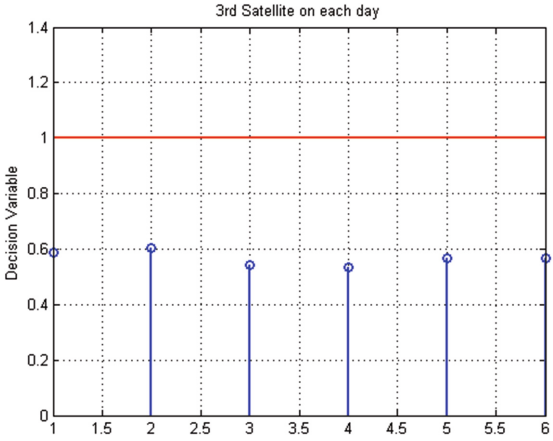


Fig. 6. Decision for 3rd satellite

From the Fig. 2 the 7th satellite is affected more due to the ionospheric variations throughout the analysed days and that the 3rd satellite is least affected. This can be justified by checking the arrangement of the satellite from the skyplot [9] as shown in Fig. 7. It describes the relative arrangement of the satellite constellation with respect to the receiver location. The signal from satellite 7 has to traverse a longer distance through the ionsophere whereas it is shorter for the 3rd satellite.

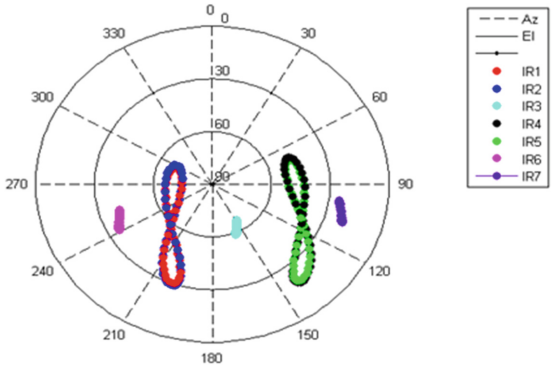


Fig. 7. Skyplot of IRNSS seven satellites [9]

4 Conclusions

Data was observed for the 2nd week of September 2016 using IRNSS receiver. The data was compiled for each satellite PRN seperately and the FAR algorithm

applied. The satellites are classified as per the threshold found using the inverse chi-squared technique. The results obtained after application of algorithm is in agreement with that of the TEC data associated with each satellite. The algorithm is able to distinguish the satellites that are affected by large variation in the ionosphere. When sufficient number of satellites are available, the satellites with higher value of decision variable can be avoided for PVT measurement. Thereby improving the accuracy in positioning.

Improved positioning system can help in various applications like navigation, disaster management, intelligent vehicle tracking and vehicle routing to avoid traffic jams using developing IoT implementations. The applications are innumerable.

References

1. Shukla, A.K., Shinghal, P., Sivaraman, M.R., Bandyopadhyay, K.: Comparative analysis of the effect of ionospheric delay on user position accuracy using single and dual frequency GPS receivers over Indian region. Indian J. Radio Space Phys. **38**, 57–61 (2009)
2. Ruparelia, S., Lineswala, P., Jagiwala, D., Desai, M.V., Shah, S.N., Dalal, U.D.: Study of L5 Band Interferences on IRNSS. Presented on International GNSS (GAGAN-IRNSS) User Meet, Bengaluru (2015)
3. Swamy, K.C.T., Sarma, A.D., Srinivas, V.S., Kumar, P.N., Rao, P.S.: Accuracy evaluation of estimated ionospheric delay of GPS signals based on Klobuchar and IRI-2007 models in low latitude region. IEEE Geosci. Remote Sens. Lett. **10**(6), 1557–1561 (2013)
4. Desai, M.V., Shah, S.N.: Ionodelay models for satellite based navigation system. Afr. J. Comput. ICT **8**(2), 25–32 (2015)
5. Paulo, O.C., Monico, J.F.G., Ferreira, L.D.D.: Application of ionospheric corrections in the equatorial region for L1 GPS users. Earth Planets Space **52**(11), 1083–1089 (2000)
6. Kowsik, J., Kumar, T.B., Mounika, G., Ratnam, D.V., Raghunath, S.: Detection of ionospheric anomalies based on FFT Averaging Ratio (FAR) algorithm. In: International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1–3. IEEE (2015)
7. Chen, Z., Guo, N., Qiu, R.C.: Demonstration of real-time spectrum sensing for cognitive radio. In: Military Communications Conference (MILCOM), pp. 323–328. IEEE (2010)
8. Raghunath, S., Ratnam, D.V.: Detection of low-latitude ionospheric irregularities from GNSS observations. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **8**(11), 5171–5176 (2015)
9. Desai, M.V., Jagiwala, D., Shah, S.N.: Impact of dilution of precision for position computation in Indian regional navigation satellite system. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 980–986. IEEE (2016)

A New Approach to Mitigate Jamming Attack in Wireless Adhoc Network Using ARC Technique

Naren Tada(✉), Tejas Patalia, and Pinal Rupani

Gujarat Technological University, Chandkheda, Ahmedabad, India
naren.tada@gmail.com, pataliatejas@rediffmail.com,
rupani.pinal@gmail.com

Abstract. Wireless Adhoc Network is a set of wireless nodes that dynamically self-organizing into a changeable topology to design the network using any preceding framework. Two possibilities are there to communicate nodes; either node can communicate directly or by forwarding network traffic through intermediate nodes in wireless adhoc network. Various types of attacks can undoubtedly be accomplished by an opponent either by passing MAC layer protocol or sending Radio Signals. Reviewing the role of wireless adversary, which victims the packets of high importance and do not follow network architecture. Attacker will make possible efforts of making users not to use network resources and fail the communication. The authors believe that detecting Jamming Attack using Reactive Jammers is quite difficult. Our Proposed approach about Global Detection of Jamming is helpful in securing other nodes from Jammer's Activity by broadcasting Jammer's UID. For mitigation, our approach named ARC (Anti-Reactive Control) Technique which shows that Jamming Attack against Reactive Jammer can be detected using decreasing PDR and RSS values and successfully mitigated by executing channel hopping. Using NS3 simulation, the attack can be identified through the decreased in performance criteria and successfully mitigated by executing channel hopping. We have analyzed the result using NS3 Wireless Jamming Module.

Keywords: MANET · Jamming attack · NS-3 jamming module
Coordinated channel switching · ARC (Anti-Reactive Control) Technique

1 Introduction

Wireless adhoc network [2, 4] is a decentralized kind of wireless network. The network referred as Adhoc due to its basic characteristics such as it does not depend on a pre-existing framework, for example routers in wired network or AP in managed (framework) wifi networks. It is set of mobile nodes that can actively self-organize into a random and short-term topology to build the network. Adhoc network are constantly enhancing towards miscellaneous attacks and achieving ubiquitous computing. The common feature of wireless medium is to share and merge with commodity nature of wireless technologies and an increasingly sophisticated user-base, permits wifi network to be easily monitored and broadcast on. In Adhoc networks, each mobile node may

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 192–204, 2018.
https://doi.org/10.1007/978-3-319-73712-6_20

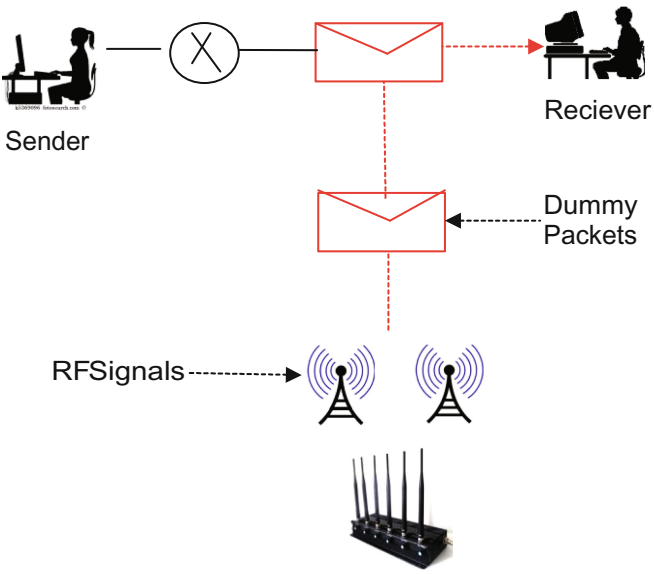
have deals directly with each other. Nodes that are not directly connected to wifi, they communicate by forwarding their traffic through relay nodes. The main benefit of Adhoc Networks is resilience, low priced and robustness. These all are qualities of adhoc network which make them well suited for military activities, emergency operations, disaster recovery, large scale community networks, and small networks. WIRELESS SECURITY [5] is the most critical attributes of Wireless communication. Mobile Adhoc Network (MANET) [2] is dynamic, independent, multi-hop network. MANET does not be in need of any fixed framework and it can be installed dynamically. Due to existence of multi-hop nature in MANET, lots of vulnerabilities present in the network. As these networks furnishing more security or more comfort zone, the issue of critical importance also come up. In MANET, different attacks such as, DDOS, Blackhole, Wormhole, Replay, Flooding, Jamming, [3] etc. have been perceived, which results in adverse effect of high level security. Since owing to fact that, Security in MANET [2, 5] is becoming challenging day by day. Attacker easily view the wireless communication between two devices and initiate simple Denial-of-Service attack against wireless network by placing distorted messages. Radio interference attacks don't seem to be available through conventional security mechanisms. An attacker will merely disregard the medium access protocol and frequently transmit on a wireless channel. On doing so, we can either intercept users to start up with legitimate MAC operations, or found packet collisions that force repeated back-offs or also jams communications.

Jamming Attack occur by continuously sending radio signals which disrupts the legitimate communication between sender and receiver. Figure 1 shows that jammer senses the communication initiated onto the wireless channel. It begins to send radio signals which injects dummy packets and receiver receives dummy packets instead of original packets send from transmitter. Attacker targets the packets of high importance. To know how jammer attacking in wireless networks and how to stay away from this jamming, researchers launch two aspects: 1. Types of existing jammers, 2. Performance Metrics. The flow of this write-up is ordered as follows: Sect. 1 incorporate an introduction to Adhoc Network, main concept and issues of MANET, Sect. 2 gives overview of Related Work, Sect. 3 gives overview Existing System, Sect. 4 shows Proposed System, Sect. 5 shows Results.

2 Related Work

2.1 Jamming Efficiency Metrics

A strong belief is there that Jammer consistently sends Radio Signals in wireless channels which results in effective blocking of channel and expected recipient might not be able to receive the message. Thus the presence of jammer in network cause interference in between legitimate communication across the wireless channel. To conclude above standards, researchers defined few metrics that apprehend the jammer's activity. Looking at the situation with one Sender (Sx) and one Receiver (Rx). Xu et al. [5] found two metrics (PSR and PDR).



Pictorial view of Jamming attack.

Fig. 1. Jamming attack

Packet Send Ratio (PSR) [5, 7]:
In this article, let us acquire that n number of packets transmitting through channel. Only m (n >= m) of these packets transmitted correctly.

$$\text{PSR} = \frac{\text{m}}{\text{n}} = \frac{\text{No. of Packets Sent}}{\text{Packets Observed to be Sent}} \tag{1}$$

Packet Delivery Ratio (PDR) [5, 7]:
Let's acquire that Rx receive m packets sent from Sx. But unfortunately only q of packets broadcast successfully to Rx. Packets proceed from CRC (Cyclic Redundancy Codes) check are referred as successful acceptance of Packets. If m = 0, then PDR be zero.

$$\text{PDR} = \frac{\text{q}}{\text{n}} = \frac{\text{Packets undergo CRC}}{\text{No. of Packets Received}} \tag{2}$$

2.2 Types of Jammer [5, 6, 8]:

Proactive Jammer
Proactive Jammers do not assure that any data communication is going on in wireless channel or not. This jammer keeps imparting Jamming Signals and disrupts the network. In case, some channel's status is ON; it initiates to send random bits onto that wireless channel. There are three types of proactive jammer: Constant, Deceptive and Random Jammer.

Constant Jammer
A constant jammer persistently producing radio signals on the wireless channel. The purpose of this type of jammer is dual: (a) to raise interference on any of the transmitting node in a way to distort its packets at the receiver (lower PDR) and (b) to form an authorized sender that (by using carrier sensing mechanism) sense the channel busy, thus preventing it from acquiring access to the channel (lower PSR).

Deceptive Jammer
Persistently dispatching normal packets instead of transmitting random bits (during the time of constant jammer). It misguides other nodes to assume that some genuine activity going on. As a consequence, they continue to exist in receiving states up to the time the jammer is turned off or dies. Alike to the constant jammer, deceptive jammer is energy ineffectual because of the constant transmission, but is straightforwardly executed.

Random Jammer
This Jammer periodically send either random bits or normal packets into network. Conflicting to the above two jammers, it targets to save energy. It constantly moving by linking two states: sleep and jamming phase. It sleeps for a certain amount of time and then comes in an operative/working mode for jamming before it goes back to a sleep state. The sleeping and jamming time periods are either fixed or random. There is a trade-off between jamming effectiveness and energy saving as it can't be jammed at the time of its sleeping phase. The ratios between both phase can be handled to regulate this trade-off between efficiency and effectiveness.

Reactive Jammer [5]:
Reactive jammers go ahead for jamming only when It discover some network activity arise on a few channel. It can distort small and large sized packets. After all it has to repeatedly watchdog the network; as reactive jammer is less energy efficient than random jammer. Upcoming are two different techniques to implement a reactive jammer.

RTS/CTS
It initiates jamming the network instantly when it observes that request-to-send packets transmitted from sender. As the attacker get aware that RTS packet transmitted in channel, attacker will distort this packet and thus receiver will not be able to send Clear-to-send (CTS) packet to sender. Until Sender don't get CTS response, it will send data to receiver and assumes that receiver is engaged with some other transmissions. This complete process will result in Jammer stay in standby position till CTS message

sent by receiver. It will jam CTS packet when transmit from receiver which will make sender not sending data and also receiver perpetually wait for data packets to receive.

DATA/ACK
This kind jams the channel by modifying the data packets or acknowledgement (ACK) response. As per the main characteristics of reactive jammer, DATA/ACK will also not do any disruption until communication start on the channel. DATA/ACK jammers corrupt the packet when it reaches to destination and till that it will suborn packet or it will be in standby position. The alteration of both packets shows re-transmissions at the sender end. Whenever information packets were not ready to receive it exactly, they need to be retransmitted. At the time when sender does not receive ACK packets, it imagines that one thing is wrong at receiver aspect, as just in case of buffer overflow, that once more ends up in re-transmission of information packets.

2.3 Types of Jamming [18, 28]:

Physical Jamming
Physical jamming during wireless network is uncomplicated but it causes different forms of DoS attack. These attacks mainly jam the channel or network by repeatedly sending jamming signals or radio frequency signals or by sending random packets. It keeps complete control over the wireless medium. This makes waste of time as each node enter into the waiting phase and need to wait till the time jammer deactivate itself and channel becomes idle to communicates.

Virtual Jamming
The usage Virtual carrier sensing mechanism done at MAC (Media Access Control) layer. To determine the presence of jammer in network, virtual jamming plays an important role. There are several benefits of MAC layer such as rival nodes; less power consumption is there compare to physical jamming. In MAC layer, the effect of Jamming initiates by attacking on the RTS/CTS frames or DATA/ACK frames.

3 Existing System [1]

Considering scenario in which system utilizing four honest nodes in the adhoc network topology. These honest nodes named as source (node 0), recipient (node 3) and remaining two trusted relay nodes (node 1 and 2). Also one Jammer Node present in the network. The base class provide strategy to detect presence of jammer using decreasing RSS and PDR values. In this system, authors presuming that there is not any of direct link from source node to destination node for legitimate communication. Existing Jamming Strategy introduced two trusted relay nodes (act as intermediate node) to transfer message to destination. The working of their strategy begins with node 0 first transmits message to both trusted relays and then forward to destination.

The work-flow starts from Physical layer to wireless module utility whose base class imparting special functions for jamming mitigation to operate. Especially for jammer

node, the strategy focuses on three different jammers': Random, Reactive and Constant. Firstly, Physical layer receive packet and forward to wireless module utility which keeps the record of packet information and calculate its RSS and PDR. After deciding on which channel jammer is activated, based on information, honest nodes request to work with different channel than one used by jammer to stay away from jamming effect.

4 Proposed System

4.1 Jamming Model

When Jammer observe that any communication initiated onto the channel, Jammer will start sending RF signal which leads to completely jammed channel. As soon as communication starts disrupting due to jamming effect, entire network nodes split into three groups named as Fully Jammed nodes, Partially Jammed Nodes and Unaffected Nodes.

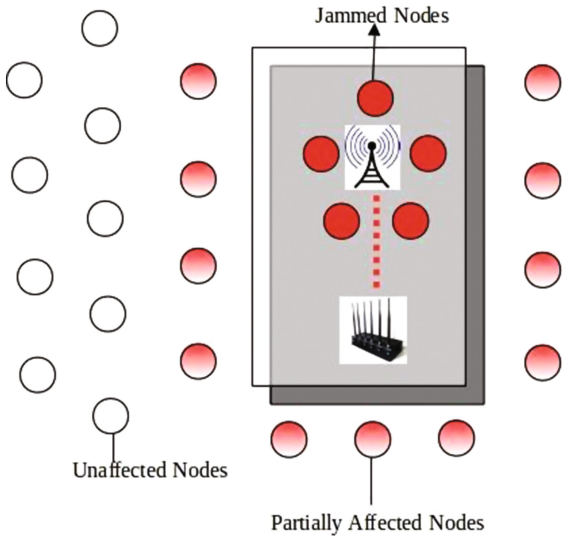


Fig. 2. Working of jamming model

As shown in Fig. 2, nodes which are nearby to jammer's position start getting affected in few seconds and ultimately it cannot receive packets from any of its neighbors. These types of nodes referred as "Fully Jammed Nodes". The area in which nodes get highly affected by jammer, declare that part as "Jammed region". Nodes which are placed at the edge of jammed region, is not completely jammed, but some part of its neighbors are jammed and referred as "Partially Jammed Nodes". This type of nodes can still reach to at least one unaffected nodes, possibly, during multi-hop nature. "Unaffected nodes" are those nodes which are placed at outermost part of the jammed region and it don't get influenced from the jamming effect.

4.2 Proposed Flowchart

See Fig. 3.

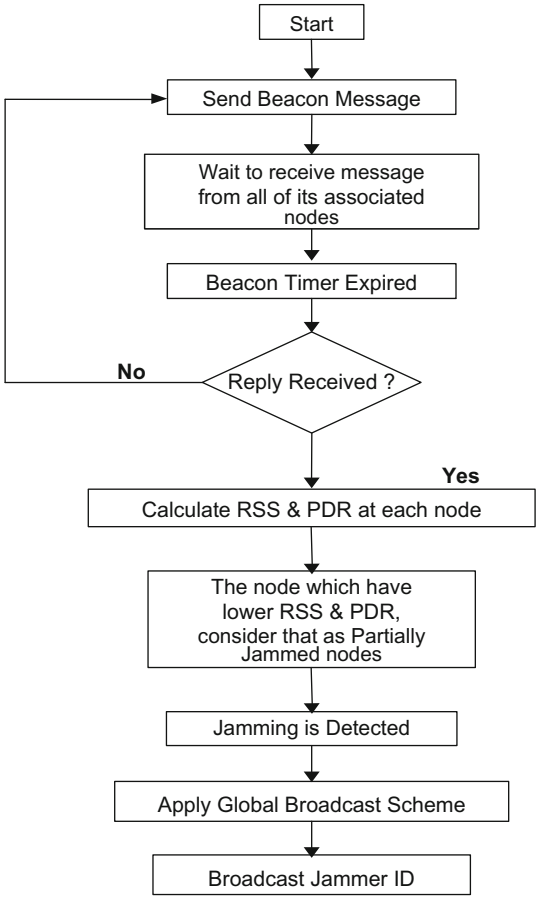


Fig. 3. Flowchart of Global Detection Scheme

4.3 Proposed Theory

Jamming Detection Intelligence

Asignificant amount of research has been devoted to study security issues as well as countermeasures to various attacks in MANET. However, there is still much research work needed to be done in this area. The aim to study is to mitigate Jamming Attack under Reactive Jammers in MANET. The proposed work is based on scheme for detecting the jamming effect in the network.

The proposed work is based on scheme for detecting the jamming effect in the network. For this, each and every node will have own unique ID (UID). In this scheme, after 10 s of simulation, the Base station will send Beacon Message to all of its associated nodes and wait for their beacon response. When response received by any of the nodes, we are calculating RSS and PDR values at each node. If the values are decreasing at some node, then we consider that node as Partially Jammed or Boundary Node. Thus from boundary nodes, we will declare their neighbor nodes as jammed node and jamming attack is thus detected. For Global Detection we are estimating the jammer's position and get its unique ID (UID). After finding the Jammer Unique ID, we are broadcasting message that "Do not receive packets from Jammer Unique ID". Thus we are securing all other nodes from Jammer's activity

Jamming Mitigation Intelligence

We proposed new approach referred as ARC (Anti-Reactive Control) Technique for Mitigation Strategy that comes up with base class which is Channel Hopping using Random Sequence Generator (RNG) Method. Alike in this strategy, on detection of jamming attack, channel hop scheme executes and each node shifts its current channel. As per proposed Global Detection Scheme, we are broadcasting Jammer's UID to each node which secure them from jammer's activity.

Channel Hopping [21]:

The graphical view of Channel hopping is shown in below figure. For channel hopping we are using Random Sequence Generator (RNG) Method if Channel Hop message is executed on detection of Jamming Attack. In this method, we applied logic to hop the channel based on automatic approach. This RNG method will return next channel number to switch the network nodes.

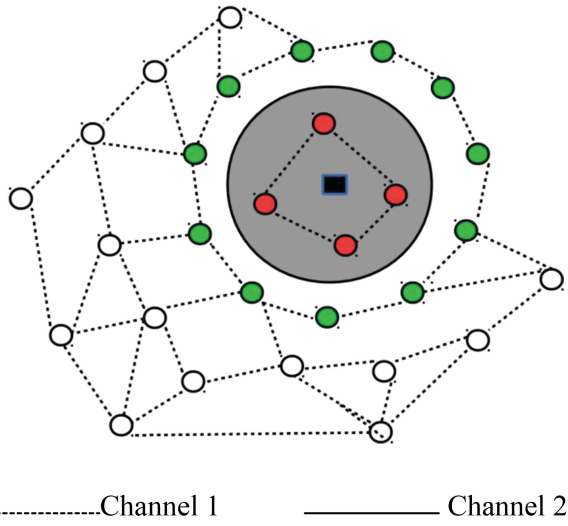


Fig. 4. Detect neighbors are not present in network (Color figure online)

Figure 4 shows that green nodes are boundary nodes of jammed area. Here nodes are realizing that their neighbors are missing from the network. After detecting that neighbors are not present in network each node will calculate its PDR and RSS value and based on results Jamming will be detected. On Detection of Jamming, RNG method will be executed and it will return next channel number to switch the communication operation.

Figure 5 shows that on execution on channel change command, each node change their channel and again start communication.

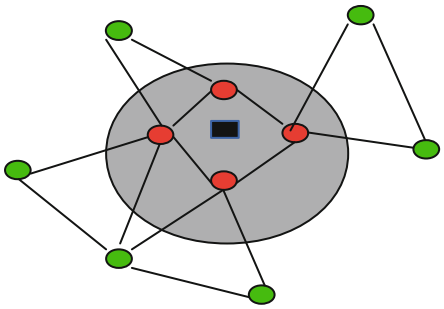


Fig. 5. Each node hops to a new communication channel

5 Results and Implementation

5.1 Implementation

Implementation of Jamming Detection and Mitigation Strategies done in NS-3 [17]. In this research we have integrated Jamming Module basically originated from https://www.nsnam.org/. To implement our strategy, we have done few modifications in the code of jamming model. All the alteration done under Reactive Jammer.

5.2 Simulation Parameters

See Table 1.

Table 1. Simulation parameters used in implementation of ARC technique

Parameter	Value
Number of packets	10000
Interval	1
Start time	0.0 s
Size of packet	100 bytes
Distance to Rx	5.0 m
Beacon port	80
Number of nodes	4 (Wifi Nodes) + 1 (Jammer Node) = 5

5.3 Results

Figure 6 shows that PDR of network. In this experiment, jammer activates node 2 and node 3 is jammed by jammer. We observed that RSS and PDR decreasing at node 2 which is boundary node. As per our results, from total simulation time of 60 s, communication stops earlier 17 s.

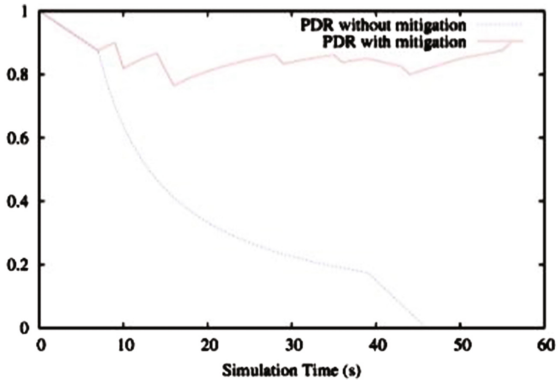


Fig. 6. PDR comparison

Figure 7. shows that RSS significantly increasing in the network as jammer's is continuously creating disturbance before implementation of ARC Technique. In 60 s of simulation RSS is not decreasing and constantly stay up to 4500 pW.

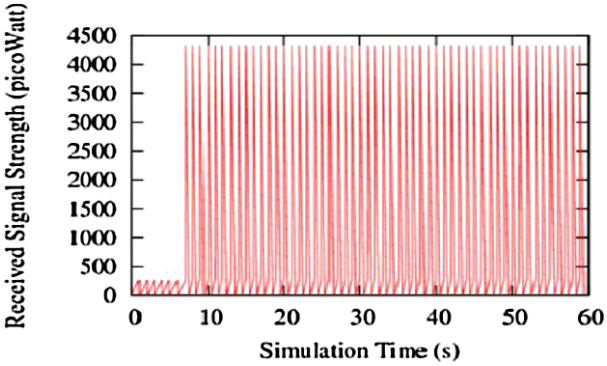


Fig. 7. RSS before mitigation

Figure 8 shows the effect of proposed mitigation strategy. It has been observed that RSS stay under pico watt.

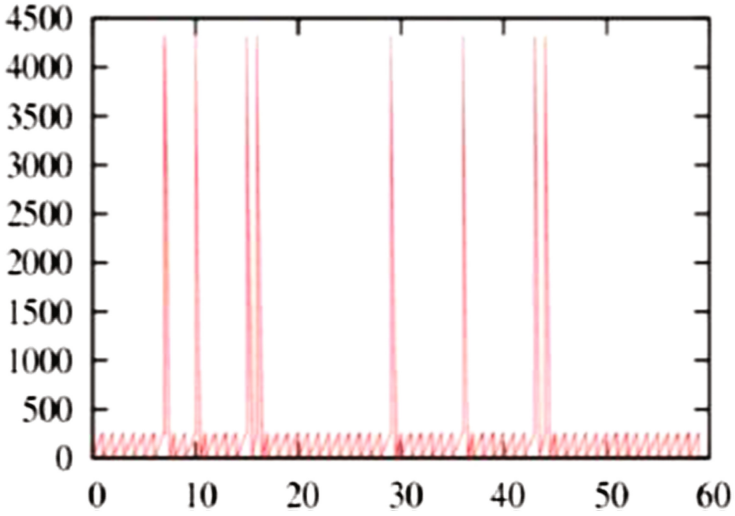


Fig. 8. RSS after mitigation

6 Conclusion and Future Scope

6.1 Conclusion

Mobile Adhoc Network (MANET) is a kind of Adhoc network with mobile, wireless nodes. Its special characteristics like open network boundary, dynamic topology and wireless communications made security highly challengeable. Jamming attack disrupts normal communication by sending continuous radio signals onto that channel.

In this research work, by studying a lot on jamming attack, we proposed new approach called ARC (Anti-Reactive Control) Technique to mitigate Jamming Attack under reactive jammers in wireless adhoc network. The proposed system will be used for Global Detection of Jamming attack and mitigating effect of jamming attack. The main advantage of our ARC Technique is "Global Broadcast Scheme", through which we are able to secure other nodes from effect of Jamming attack from Reactive Jammers. Other main advantage is Channel Hopping using Random Sequence Generator method.

6.2 Future Scope

Overall Technique works best and fulfills its objectives but this technique works with only one Jammer. The future target is to introduce mitigation scheme by placing multiple reactive jammer's in the network. We are hopping each node in the next channel but for future studies we should also study that only jammed node change its channel of communication.

References

1. Kushardianto, N.C., Kusnanto, Y., Syafruizal, E., Tohari, A.H.: The effect of jamming attack detection and mitigation on energy power consumption (Case study IEEE 802.11 wireless adhoc network). Jurnal Teknologi **77**, 39–46 (2015)
2. Kumari, S., Sanduja, M.R.: Detection of jamming attack in Mobile Adhoc Network. IJSETR **5**(6), June 2016
3. Kannhavong, B., Nakayama, H., Nemoto, Y., Kato, N., Jamalipour, A.: A survey of routing attacks in Mobile Adhoc Network. IEEE Wireless Commun. October 2007. Tohoku University, Sydney
4. Rubinstein, M.G., Moraes, I.M., Campista, M.E.M., Costa, L.H.M.K., Duarte, O.C.M.B.: A survey on wireless ad hoc networks. In: Pujolle, G. (ed.) MWCN 2006. ITIFIP, vol. 211, pp. 1–33. Springer, Boston (2006). https://doi.org/10.1007/978-0-387-34736-3_1
5. Pelechrinis, K., Iliofotou, M., Krishnamurthy, S.V.: Denial of service attacks in wireless networks: the case of jammers. IEEE Commun. Surv. Tutor. **13**(2), 245–257 (2011). Second Quarter
6. Grover, K., Lim, A., Yang, Q.: Jamming and anti-jamming techniques in wireless networks: a survey. Int. J. Adhoc Ubiquit. Comput. **17**, 197–215 (2017)
7. Xu, W., Ma, K., Trappe, W., Zhang, Y.: Jamming sensor network: attack and defense strategies. IEEE Network **20**, 41–47 (2006)
8. Xu, W., Trappe, W., Zhang, Y., Wood, T.: The feasibility of launching and detecting jamming attacks in wireless networks. Wireless Information Network Laboratory (WINLAB) Rutgers University, 73 Brett Rd., Piscataway, NJ 08854, 25–27 March 2005
9. Naren, T., Tejas, P., Chirag, P.: Trust appraisal based neighbour defense secure routing to mitigate various attacks in most vulnerable wireless ad hoc network. In: Satapathy, S.C.C., Das, S. (eds.) Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1. SIST, vol. 50, pp. 323–332. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30933-0_33
10. Popli, P., Raj, P.: Mitigation of jamming attack in Mobile Adhoc Network. IJIRCCE **4**(6) (2016)
11. Popli, P., Raj, P.: Securing MANET by eliminating jamming attack through mechanism. IJSETR **5**(9), September 2016
12. Liu, H., Liu, Z., Chen, Y., Xu, W.: Determining the position of a jammer using a virtual-force iterative approach. Wireless Netw. **17**, 531–547 (2010). Springer Publication
13. Misra, S., Dhurandher, S.K., Rayankula, A., Agrawal, D.: Using honeynodes for defense against jamming attacks in wireless infrastructure-based network. Comput. Electr. Eng. **36**, 367–382 (2009). Elsevier
14. Vijayakumar, K.P., Ganeshkumar, P., Anandaraj, M.: Jamming detection system in wireless sensor networks. IJARCET **3**(4), April 2014
15. Liu, H., Xu, W., Chen, Y., Liu, Z.: Localizing jammers in wireless network. Dept of ECE, Stevens Institute of Tech. Castle Point on Husdon, Hoboken, NJ 07030 and Dept of CSE, Uni. Of South Carolina, Columbia, SC 29208
16. Popli, P., Raj, P.: Effect of jamming attack in Mobile Adhoc Environment. IJSETR **5**(5), May 2016
17. Khosla, H., Kaur, R.: Jamming attack detection and isolation to increase efficiency of the network in Mobile Ad-Hoc Network. IJRET **2**(4), July 2015
18. Liu, Z., Liu, H., Xu, W., Chen, Y.: Exploiting jamming - caused neighbor changes for jammer localization. IEEE Trans. Parallel Distrib. Syst. **23**(3), 547–555 (2012)

19. Xu, W., Wood, T., Trappe, W., Zhang, Y.: Channel surfing and spatial retreats: defense against wireless denial of service. In: Proceedings of ACM Wireless Security, 1 October 2004
20. Ajana, J., Helen, K.J.: Mitigating inside jammers in MANET using localized detection scheme. IJESI **2**(7), 13–19 (2013)
21. Zhang, R., Sun, J., Zhang, Y., Huang, X.: Jamming-resilient secure neighbor discovery in Mobile Ad Hoc Networks. IEEE Trans. Wireless Commun. **14**, 5588–5601 (2015)
22. Thuente, D., Acharya, M.: Intelligent jamming in wireless networks with applications to 802.11b and other networks. IEEE
23. Mpitziopoulos, A., Gavalas, D., Konstantopoulos, C., Pantziou, G.: A survey on jamming attacks and countermeasures in WSNs. IEEE Commun. Surv. Tutor. **11**(4) (2009). Fourth quarter
24. Thamilarasu, G., Mishra, S., Sridhar, R.: A cross-layer approach to detect jamming attacks in wireless ad hoc networks. IEEE
25. Ben-Othman, J., Hamieh, A.: Defending method against jamming attack in wireless ad hoc networks. IEEE, 20–23 October 2009
26. Hamieh, A., Ben-Othman, J.: Detection of jamming attacks in wireless ad hoc networks using error distribution. IEEE (2009)
27. Ashraf, Q.M., Habaebi, M.H., Islam, M.R.: Jammer localization using wireless devices with mitigation by self-configuration. PLoS One (2016)
28. Chaturvedi, P., Gupta, K.: Detection and prevention of various types of jamming attacks in wireless networks. IJCNWC **3**(2), 2250–3501 (2013)
29. Kopena, J.: Wireless Jamming Model. https://www.nsnam.org/wiki/Wireless_jamming_model
30. NS3 Official Site: https://www.nsnam.org/

Optimize Spectrum Allocation in Cognitive Radio Network

Nidhi Patel[1(✉)], Ketki Pathak[2], and Rahul Patel[1]

[1] Electronics and Communication Department, Dr. S. & S. S. Ghandhy Engineering College, Surat, India
nidhipmanoj@yahoo.com, rmpgec@gmail.com
[2] Electronics and Communication Department, S.C.E.T., Surat, India
ketki.joshi@scet.ac.in

Abstract. With rapid evolution in wireless devices increases the demand for radio spectrum. To solve spectrum underutilization problem cognitive radio technology is introduced. Cognitive radio technology is next generation technology which allows non-licensed user to use electromagnetic spectrum without interfering licensed user. To use white space in radio spectrum one should sense the spectrum perfectly. Once sensing is done, the distribution of the spectrum among the secondary user is also challenging task. Optimizing is the process to find best solution among the available solutions. Radio environment is random in nature. Due to fast convergence property of the genetic algorithm can use to find optimal solution for spectrum allocation problem to maximizing spectral utilization. Problem is modeled as Multi Objective Problem (MOP), considering that function as fitness function and evaluating the best allocation among all. Firstly defining target objective function that is minimizing Bit Error Rate (BER), maximizing throughput and minimizing power, then using aggregate sum approach, it converts all single objective function into one MOP. Than mathematically applying the fitness function in software so we get graphical representation. We have check convergence of algorithm first. Than we simulate result for single channel and multichannel performance. By observation of graphical parameter we have simulate results for real scenario and get optimum parameter for given situation.

Keywords: Cognitive radio · Spectrum allocation · Primary user
Secondary user · Dynamic spectrum access · Genetic algorithm

1 Introduction

Over the last two decades the use of wireless devices rapidly, which increase the shortage of spectrum resource. Generally Electro Magnetic (EM) spectrum is regulated by governmental bodies like Federal Communication Commission (FCC). According to FCC 70% of the spectrum stays unused most of the time [1]. Figure 1 shows the distribution of signal strength among large band of frequency. The fixed spectrum allocation was serve well in the past but increases the wireless devices the fixed spectrum policy cannot work well.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 205–214, 2018.
https://doi.org/10.1007/978-3-319-73712-6_21

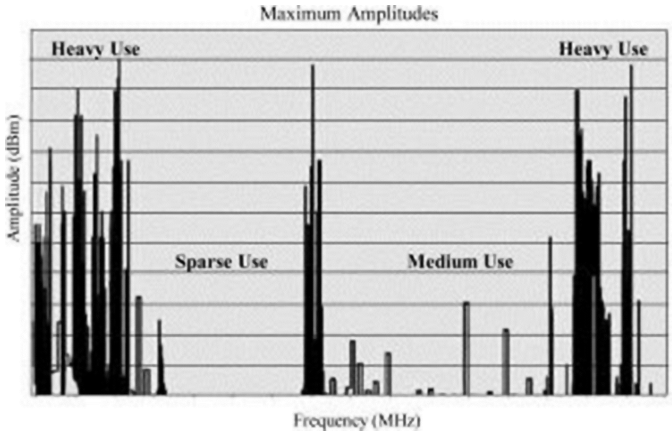


Fig. 1. Spectrum usage [1]

To solve the underutilization problem Cognitive Radio (CR) works well. CR is capable of transmitting in the licensed band without causing harmful interference to the Primary User (PU). "Cognitive Radio" can be defined as the radio that can change its transmission parameter based on interaction with environment [2]. There are two characteristics of the CR that are cognitive capability and cognitive re-configurability [3]. Cognitive capability defines as the ability to capture or sense the information from radio environment. In more specific word CR allows Secondary User (SU) to detect, which portion of the EM band is not in used. Select the best available channel according the SU's requirement, there are so many SUs are accessing the band so coordinating the channel access among all SUs and vacant the channel when primary user wants to use the channel. All functionality related to each other is described in Fig. 2.

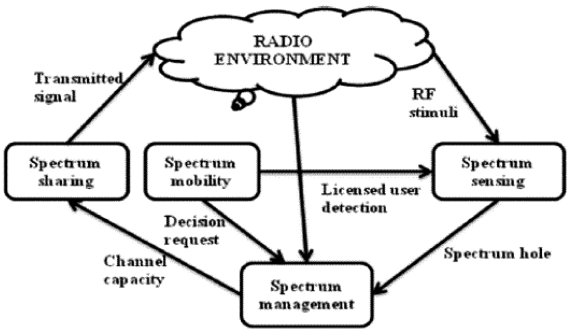


Fig. 2. Flow diagram of CRN

The main function of cognitive radio is summarized as follows:

Spectrum sensing: A CR can only use the vacant band of the radio spectrum, which is not currently used by primary user. Therefor CR should monitor the available spectrum, capture their information and detect the spectrum hole [3]. Spectrum management:

based on spectrum availability, the channel is allocated to SU according to its criteria [3]. Spectrum sharing: since there are multiple users are accessing the spectrum hole, CR access should be coordinate in order to avoid the collision between them [3]. Spectrum mobility: when primary user wants to use the channel the SU have to vacant that channel to avoid interference with PU and switch to other channel [3].

2 Cognitive Radio Operating Parameters

To developing cognitive radio control system we must define some control parameters to the system. The quality and quantity of control parameter decide how accurate system performance is.

2.1 Transmission Parameter

CR takes advantage of control parameters. These control parameters are input to the fitness function. Fitness function is the scalars score how well the fitness value met the optimum value. For generation of fitness function we must have some list of parameters. These transmission parameters specifically used as control parameters. We used three control parameter are transmission parameter, modulation scheme and modulation index to construct the fitness function [4, 5].

2.2 Environmental Parameter

Environmental measures inform system about surrounding. They may be internal or external information. Internal parameters are regarding to internal state of the mobile device e.g. battery life. Environmental parameters are classified into two categories that are primary parameters and triggering parameters. Primary parameters are directly put into fitness function. So the effect of primary parameters observed from the fitness function i.e. noise power and interference power. Triggering parameters are supervised by the system that is battery life. If system found low battery it automatically switch low power weighing scenario [4, 5].

2.3 Performance Objective

In wireless communication there are lots of desirable performance objectives to achieve desire Quality of Service (QoS) of mobile device. Here we choose three performance objectives that are minimizing BER, minimizing power and maximizing throughput. There is no particular method to optimize all objective function simultaneously but there is always trade off.

Minimizing BER. Fitness function of BER required three control parameters, those are Transmission power, modulation index and modulation type. Environmental parameters noise power. Specifically we used m-ary PSK modulation type. The formula for probability of BER is as follows [4, 5, 7, 9].

$$P_{be} = \frac{2}{log2(m)} Q\left(\sqrt{2 * log2(m) * \gamma} * sin\frac{\pi}{m}\right) \tag{1}$$

Objective function for minimizing BER for single carrier system is,

$$f\min_{ber} = 1 - \frac{\log 0.5 - \log P_{be}}{\log 0.5 - \log 10^{\wedge -6}} \tag{2}$$

For multicarrier objective function we have,

$$f\min_{ber} = 1 - \frac{\log 0.5}{\log\left(\overline{P_{be}}\right)} \tag{3}$$

Where, $\overline{P_{be}}$ is the average BER over N independent subcarriers.

Maximizing Throughput. Control parameters are modulation index and modulation scheme are utilize by objective function [4, 5, 7].

$$f\max_{throughput} = 1 - \frac{\sum \log mi}{N \log m_{max}} \tag{4}$$

Where mi is the number of bits per symbol m_{max} is the maximum modulation index and N is the number of sub carriers.

Minimizing Power. Fitness function consist three control parameters are transmission power, modulation index and modulation type [4, 5, 7].

$$f\min_{power} = \sum_{i=0}^{N} 1 - \left(\frac{P_i}{P_{max}}\right) \tag{5}$$

P_i is the transmitting power, N is the number of subcarriers and P_{max} is the maximum value of the power transmitted for any subcarrier.

2.4 Multi Objective Goals [9]

We propose sum aggregate approach to combine single objective functions to one objective function as fitness function. Each objective function is multiplied by weight value and sum of each weight together gives us single scalar value that is 1.

$$f_{multicarrier} = w_1 * f\min_{ber} + w_2 * f\max_{throughput} + w_3 f\min_{power} \tag{6}$$

2.5 Genetic Approach [8]

In general solution to any problem represent by binary string. If these strings allow going under binary growth, good strings are split and poorer fitness string are die out. This decision is taken by the fitness function. Genetic algorithm possesses these characteristics. Flow diagram of genetic algorithm is shown in Fig. 3. GA is implemented using four basic steps are initialization, fitness measure, reproduction of new population and stopping criteria.

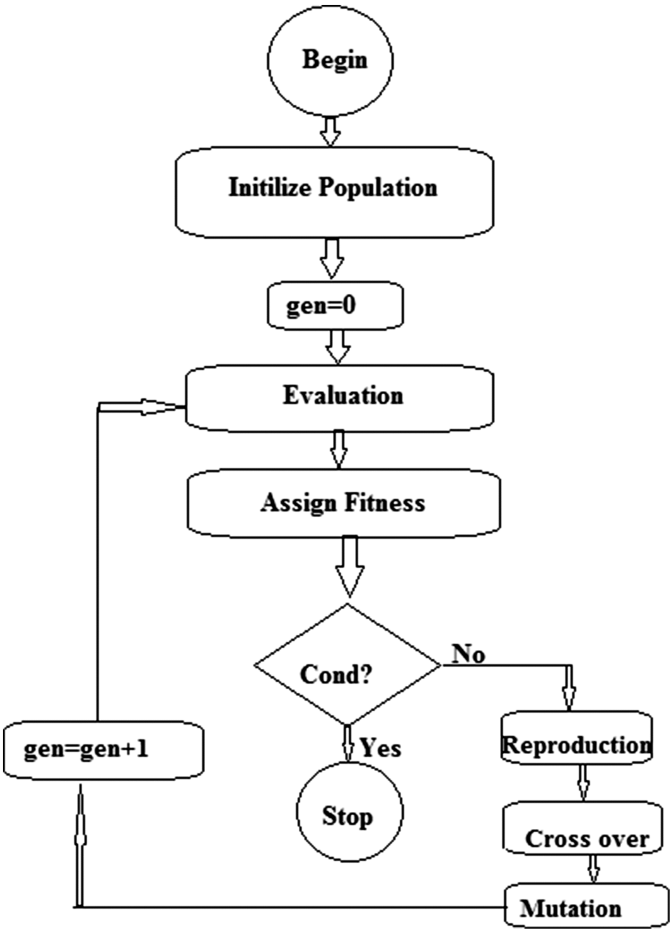


Fig. 3. Flow diagram of genetic algorithm [4]

Initialization: A random initial population of 'n' (number of initial population) chromosomes is generated. This population contains the available solutions for the specified problem.

Fitness Measure: Evaluation of the fitness of an initial population's chromosomes.

Construction of New Population: Try the following steps to reproduce, until the production of the next generation completes.

Selection: A selection of chromosomes will be done in a way such that these chromosomes have the better level of fitness in the current available population.
Crossover: The crossover is done to make new individuals for the incoming generation. So with the defined probability of crossover, selected chromosomes reproduce to form new individuals.
Mutation: The new created individual will be mutated at a definite point.

Stopping Criteria: The process is repeated with all the above mentioned steps until a desired optimum solution is obtained or a set of maximum numbers of the population are generated. This stage is termination stage. The genetic algorithm process detailed previously continues until a termination condition has been reached. Common termination conditions includes, A solution that satisfies minimum criteria is found, A fixed number of generations is reached, A specified computation time is reached and The fitness scores have plateaued such that successive generations show no improvement.

To implement the GA there are still several factors to consider, like creation of chromosomes, types of encoding used to perform the genetic algorithms, selection of the optimum chromosomes, and different criterion such as defining the fitness measure.

3 Simulation and Results

GA simulation converges very fast to the optimal value. Once it reaches nearer optimum, if we increase number of iteration which increases the processing time with little

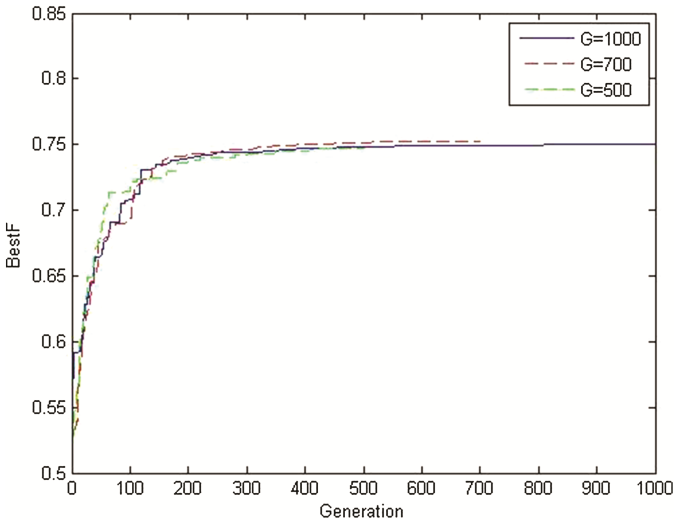


Fig. 4. Fitness convergence plot for varying number of generation

improvement in the fitness. Processing time is critical factor in wireless communication. Getting optimum iteration is also challenging task (Fig. 4).

Begin with single objective function, minimize BER performance objective results. Figure 5 shows a standard fitness convergence graph obtained from the GA system. This figure shows the results from varying channels in the system. It can be seen that a system with a single channel converges much faster than the system with 2 channels as well as the processing time needed to calculate the fitness over a 2 channel system. To highlight the effect of the increasing number of channels in the system, Table 1 shows the optimal generation where the highest fitness was found for each system. Again, for a single channel system, the system is able to find the best value much earlier than the system with 2 channels.

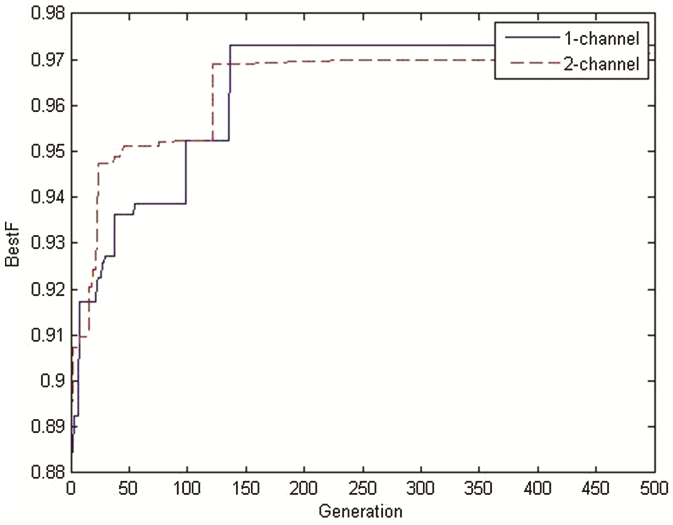


Fig. 5. Fitness convergence graph for varying number of channel

Table 1. Comparative analysis of varying number of channel

Channel	Iteration (G)	Time elapsed to run code (s)	Best fitness
1	500	29.2169	0.9796
2	500	42.5030	0.9766
1	600	37.2530	0.9850
2	600	51.4660	0.9114

3.1 Multi Objective Performances

Simulation setup for multi-objective function, we have power varying from 0.01 to 2.56 normalized values for 8 bit chromosome. In which 6 bits used for power allocation and 2 bit for assigning modulation index, varying from 2, 4, 8 and 16. Now a days higher

modulation index also used for practical purpose so we can increase one bit in chromosome. That gives us eight different possibilities for modulation index. So the system performance is enhanced. The comparative analysis of 8 bit chromosome and 9 bit chromosome is shown in Figs. 6 and 7 as length of chromosome increases we have more combination of chromosome.

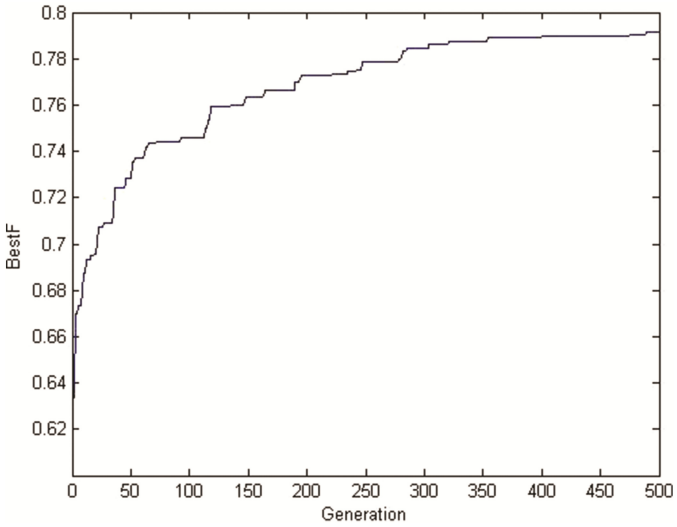


Fig. 6. Fitness convergence curve for 8 bit chromosome

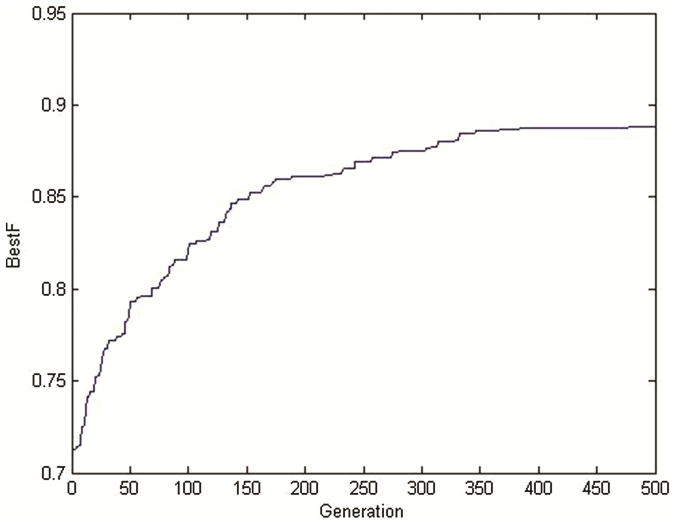


Fig. 7. Fitness convergence curve for 9 bit chromosome

Each subcarrier has a random channel attenuation N, using this value and the vector weights, the GA has optimized the transmission parameters for different mode of operation. Weights to different mode as shown in Table 2. Results are shown for voice application (Fig. 8).

Table 2. The weight value of three different modes

	Voice	Download	Low power
W1 (BER)	0.75	0.10	0.15
W2 (power)	0.10	0.15	0.75
W3 (Throughput)	0.15	0.75	0.10
Best fitness	0.7673	0.9297	0.8736

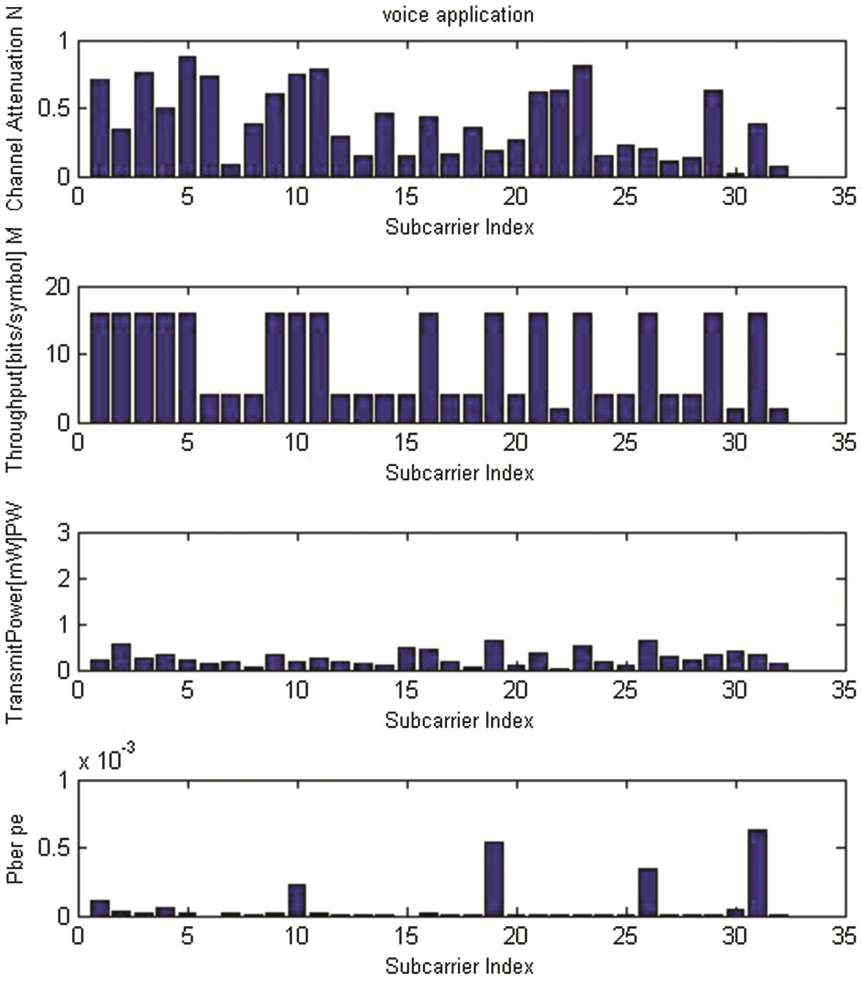


Fig. 8. Optimum solution set for voice application

4 Summary

This paper introduced an implementation of a multicarrier cognitive radio that uses a genetic Algorithm as the decision method. We have introduced several fitness functions that are used to score how well a parameter set match the given objectives. A 32 subcarrier system was then simulated using three separate scenarios. The results of these simulations proved that the fitness functions steer the evolution of the GA in the correct direction to optimize the given objectives for each scenario. An increase in the initial population will decrease the chances of premature convergence of the algorithm, but the execution time will increase accordingly. As time play very important role, single channel approach can be used to minimize time.

References

1. FCC: FCC. 03-322-notice of proposed rule making and order. Technical report, Federal Communications Commission, 30 December 2003
2. Muchandi, N., Khanai, R.: Cognitive radio spectrum sensing: a survey. In: International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE (2016)
3. Ghosh, G., Das, P., Chatterjee, S.: Simulation and analysis of cognitive radio system using Matlab. Int. J. Next-Gener. Netw. **6**(2) (2014). 31. K. Elissa, "Title of paper if known," unpublished
4. Newman, T.R., et al.: Population adaptation for genetic algorithm-based cognitive radios. Mob. Netw. Appl. **13**(5), 442–451 (2008)
5. Varade, P.S., Ravinder, Y.: Optimal spectrum allocation in cognitive radio using genetic algorithm. In: 2014 Annual IEEE India Conference (INDICON). IEEE (2014)
6. Hamza, A.S., Elghoneimy, M.M.: On the effectiveness of using genetic algorithm for spectrum allocation in cognitive radio networks. In: 2010 High-Capacity Optical Networks and Enabling Technologies (HONET). IEEE (2010)
7. Pradhan, P.M., Panda, G.: Pareto optimization of cognitive radio parameters using multiobjective evolutionary algorithms and fuzzy decision making. Swarm Evol. Comput. **7**, 7–20 (2012)
8. Pradhan, P.M., Panda, G.: Comparative performance analysis of evolutionary algorithm based parameter optimization in cognitive radio engine: a survey. Ad Hoc Netw. **17**, 129–146 (2014)
9. El-Saleh, A.A., Ismail, M., Ali, M.: Pragmatic trellis coded modulation for adaptive multi-objective genetic algorithm-based cognitive radio systems. In: 2010 16th Asia-Pacific Conference on Communications (APCC). IEEE (2010)

Activity Based Resource Allocation in IoT for Disaster Management

J. Sathish Kumar(✉), Mukesh A. Zaveri, and Meghavi Choksi

Computer Engineering Department, SVNIT, Surat, India
{ds14co001,mazaveri}@coed.svnit.ac.in, meghavichoksi@gmail.com

Abstract. Efficient utilization of resources during disasters is a major and non-trivial problem. Improper resource allocations is due to lacking in knowledge of activity priorities. Due to disaster, in a major instances, communication networks are ruined. In this regard, Internet of Things (IoT) helps to a great extent in establishment of dynamic network for communication. Further, priority based stable matching algorithm is used for allocation of resources for the corresponding activities. This approach determines for maximum utilization of resources with complete accomplishment of activities efficiently. Also, we evaluated our approach with execution time and fairness of resource allocation for utility.

Keywords: Resource allocation · Disaster management
Internet of Things · Graph theory · Stable matching

1 Introduction

A recent survey conducted by United Nations Office for Disaster Risk Reduction (UNISDR), among the top five disaster hit countries India ranked third place [1]. During 2011–2015, 38 million people were affected with different types of disasters and total of 29 million dollars of economic damage occurred [2]. One of the major reasons for such a great loss is unavailability of real time network communication, lack of identifying the activities and improper utilization of resources. Internet of Things (IoT) is an emerging technological concept that uses rapid communications using Internet. It enables the devices to communicate any time, any place and any where [3,4]. Therefore, using IoT helps to formation of dynamic IP enabled network communication.

In the available literature of disaster management [1,2,5,6,9], majorly the challenges are divided the four phases. Namely, the mitigation [15] phase and preparedness phase that is the period before the disaster. In the mitigation phase various issues like public education, infrastructure improvement and critical infrastructure protection, information campaigns were addressed. Community preparedness, volunteer management, material management and emergency response plan were dealt in the preparedness phase. Likewise in the response phase, resource allocation, situation awareness, victim management and plan

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 215–224, 2018.
https://doi.org/10.1007/978-3-319-73712-6_22

implementation, call_take_dispatch and etc. were covered because this period plays an important role during and shortly after the disaster. Whereas, in the recovery phase where the period long time after the disaster, issues like damage assessment, procurement, public information and insurance claim were considered. The glance of the four phases is as shown in the Fig. 1.

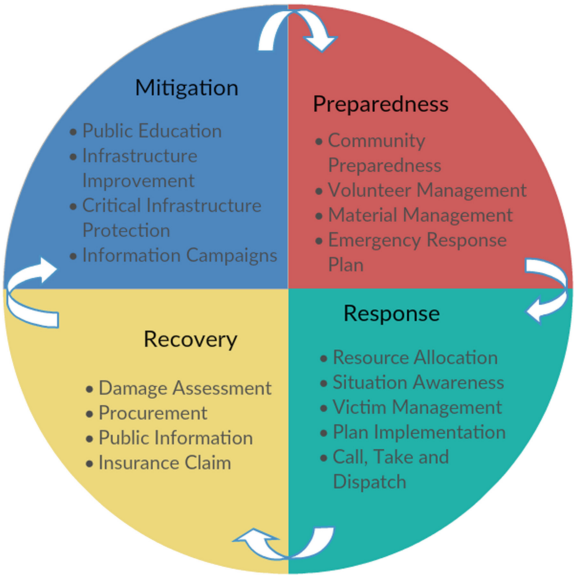


Fig. 1. Different phases of disaster management

Subsequent to the disaster, resource scheduling in the response phase is crucial that must dealt urgent based on identifying and managing tasks. Resource scheduling can be accomplished in the field level and administration level. However, on field addressing of resource scheduling is critical that directly involves in rescue, retrieve and saving lives. Hence, by studying these situations, we address the critical response for resource scheduling using Stable Matching Algorithm. Although, in the available literature for resource scheduling is applied for stable matching for paring the people for marriage [12,13] results in the safe allocation. Likewise, we further extend this approach with appropriate modifications for resource scheduling in a disaster management that results in the safe scheduling that leads to stable allocation.

Also, clustering of the devices and network, for efficient connectivity and communications in IoT is proposed in the in [8,10]. Identification of the places during disasters is handled using localization approached using IoT that is proposed in [11]. Assuming that dynamic network has been established immediately the disasters, communications in the network can be dealt efficiently real time using IoT. However, scheduling the available resources for different activity for

disaster management is an non-trivial problem that we are addressing in this paper. Also, we carried out the experimental simulation results with execution time, stability in the safe schedule and fairness for the utilization of resources.

Rest of the paper is organized as follows. The problem description is detailed in Sect. 2. Resource scheduling algorithm and corresponding complexity analysis are described in Sect. 3. In Sect. 4, simulation results are presented and conclusion in Sect. 5.

2 Problem Description

In this section, the problem description is described with appropriate notations that represent the resource allocation for the accomplishment of activities and assumptions.

Let us assume, we have m activities and n resources. The activities can be represented as a which can further defined as a set of sub activities, say $a_1, a_2...a_i..a_m$. Likewise, the resources are represented as r, that can further divided into set of resources defined as $r_1, r_2...r_j..r_n$. Now, the allocation is represented as a graph $G = (V, E)$, where V is a set of vertices which indicates the activities and resources, $E \subseteq \{\{a, r\} | a, r \in V, a \neq r\}$ defines the potential allocation edges. A state is a allocation $S \subseteq E$ such that for each $r \in V$, we have $|\{e|e \in S, r \in e\}| \leq 1$. An edge $e = \{a, r\} \in S$ provides utilities $l_a(e) = l_r(e) > 0$ for a and r respectively. If for every $e \in E$ we have some $l_a(e) = l_r(e) = l(e) > 0$, then it is correlated preferences. If no explicit values are given, we will assume that each vertices has an order of priorities over its possible allocation because for every vertex the utility of allocating edges is given according to their priorities. Then it is called as general preferences. In general preferences, the priority is allowed to be an incomplete list or to have ties. But, we define $P(S, a)$ to be $l_a(e)$ if $a \in e \in S$ and 0 otherwise. A blocking pair for allocation S is a pair of vertices $a, r \notin S$ such that each vertex a and r is either unallocated or strictly prefers the other over its current allocation. A stable allocation S is a allocation without blocking pair.

For instance, let us assume, we have three activities say $a1, a2, a3$ need to be addressed during disaster. Also, let us assume we have three resources say $r1, r2, r3$ are available. An illustrative example is depicted as shown in the Figs. 2 and 3. Each activity can be assigned priorities to utilize the resources to accomplish and vice-verse for the resources to address the activities. Now, $a3$ priorities to utilize the resources is in the order of $r1, r2$ and $r3$. Likewise, $a2$ priorities are in the order of $r2, r3$ and $r1$ and $a1$ needs $r1, r2$ and $r3$. Similarly the resource priorities for $r3$ is in the order of $a1, a2$ and $a3$, for $r2$ is $a1, a3$ and $a2$ and $r1$ is $a2, a3$ and $a1$.

To be precise, in the context of disaster, the activities can be classified as to established a communication network, to provide medical treatment to the critical, rescue and recovery. Suppose, these activities should be addressed in all the disaster places and to accomplish them, resources such as military force, fire engines, volunteers, ambulance and medical help are required. However, the

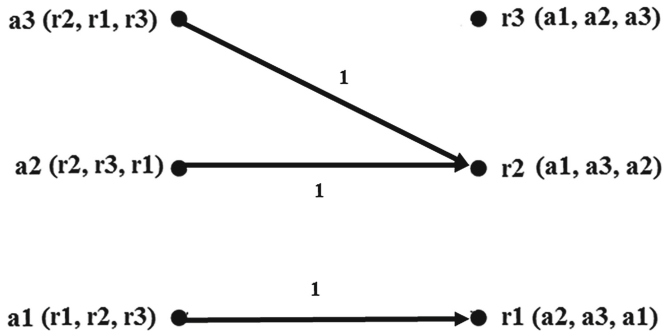


Fig. 2. Example graph with primary activity priorities

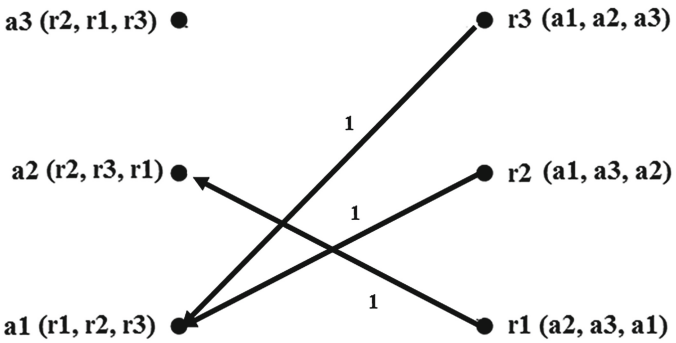


Fig. 3. Example graph with primary resource priorities

resources priorities are assigned by considering many factors such as the distance between the disaster place and resources, traffic considerations, road maintenance etc. Likewise for the activities different priorities are needed. Now, Figs. 2 and 3 depicts the demand of the same resources for different activities i.e., $a1, a2$ demands $r2$ and utility of the different resources to the same activity i.e., $r3, r2$ by $a1$ in Fig. 3. Also, few resources are not utilized properly i.e., $r3$ in Fig. 2. Hence, these kind of improper allocation leads to loss of many lives instead of saving them. Hence, to overcome them, the stable allocation of activities and resources, provided the priorities are crucial. In this regard, we propose the resource allocation algorithm that address this problem and brings the stability in the allocation which is detailed in the next section.

Since we are considering the IoT environment, we are assuming the resources has IP connectivity that by default enables to communicate any time and any where at ay place. Also, since the stable marriage allocation approach works with equal number of pairs, we assume that using clustering approach, grouping of sub activities into activities and grouping of resources can be addressed efficiently.

Further, the priorities of the activities and resources is completely depends on the context of the disaster problem. In this paper, we assume that the priorities are already defined such that we completely focused and determined to address the resource allocation efficiently. Considering these above assumptions, we propose the resource allocation algorithm that is suitable for IoT environment of addressing the resource allocation during critical disaster times.

3 Resource Allocation Algorithm

The proposed algorithm is devised in such a way that the stability in the allocation is determined efficiently. Proper utilization of all resources and proper attention to take care of all activities are considered. Having the knowledge of priorities in all the activities and resources, we utilized the stable matching concept for allocation [12,13]. In Figs. 2 and 3, since $a3$ and $a2$ are requesting the same resources $r2$, but in $r2$, the priority is given to $a1$ which leads to unstable allocation. Likewise, since $r3$ and $r2$ are requesting to be utilized by the same activity $a1$, but $a1$, needs resource $r1$ which leads to again unstable allocation.

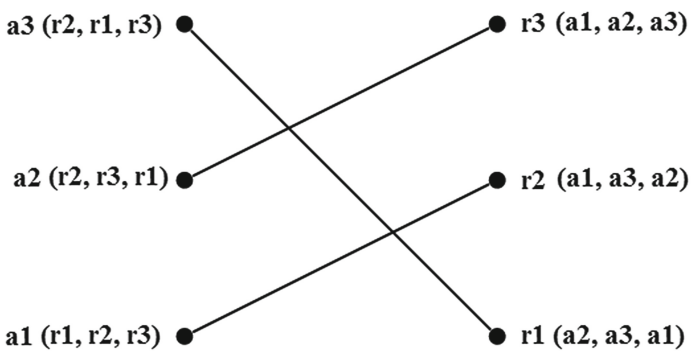


Fig. 4. Example of stable allocation of activities and resources using graph

Hence, by considering the next order priorities of both activities and resources, the allocation is been carried out in such a way that all the activities and resources got paired and none of them are left. As shown in Fig. 4, activity $a3$ is allotted with $r1$, $a2$ is allotted with $r3$ and $a1$ is allotted with $r2$ which brings the stability in the allocation. Also, complete utilization of resources and

entire activities are addressed. The corresponding algorithm is devised and as shown in the pseudo code.

Algorithm 1. Resource Allocation Algorithm

```
   Data: resources and activities with priorities
   Result: activities with allocated resources
 1 while there are still free resources and free activities do
 2 |   Let r be the first resource in the list of free resources ;
 3 |   Let a be the highest-ranked activity on r's preference list and r will
   |   try to allocate with a ;
 4 |   Let r' be the resource with activity a is currently allotted. // (r' can
   |   be -1 or some other null value if a is free);
 5 |   if a is free then
 6 |   |   r gets allotted with a ;
 7 |   |   r gets removed from the list of free resources ;
 8 |   end
 9 |   else
10 |   |   a is currently allotted to a different resource r' ;
11 |   |   if a prefers r' to r then
12 |   |   |   r stays free //(don't alter the allocation) ;
13 |   |   end
14 |   |   else
15 |   |   |   //a prefers r to r' ;
16 |   |   |   r and a get allocated with each other ;
17 |   |   |   r gets removed from the list of free resources ;
18 |   |   |   r' get added to the list of free resources ;
19 |   |   end
20 |   end
21 |   Update the next activity choice for r (even if r is no longer free) ;
22 end
23 return allocation between resources and activity pairs. ;
```

It is important to analyze the proposed algorithm in terms of computational complexity for critical time analysis and response. The proposed algorithm time complexity is $O(mn)$. There are at most mn possible allocations between activities and resources. So there at most $m \times n$ iterations. To maintain this $O(mn)$ time complexity, each iteration must therefore be of constant time due to knowledge of priorities. However, the brute force algorithm takes $O((m+n)/2!)$ since it goes through each enumeration to verify whether the allocation is stable or not. For m activities and n resources, the number of enumerations is ${}^{m}P_{n}$ which is equal to $((m+n)/2)!$.

4 Simulation Results

The simulation results for allocation of activities and resources stability is evaluated in terms execution time and fairness in the allocation resources that determines the utilization.

By First Come First Serve (FCFS) approach which works in the fashion of brute force and our proposed approach with respect to execution of time with different number of pairs are compared and is shown in Fig. 5. The proposed approach out performs the FCFS in terms of bringing stability in the allocation of resources and activities. Although till 8 pairs of resources and activities, both approaches have same execution time but as number of pairs increases the FCFS execution time rapidly grow exponentially. For larger inputs like 160, FCFS is unable execute the allocation for the resources and activities but whereas our proposed approach gives linear results even at the 1280 pairs.

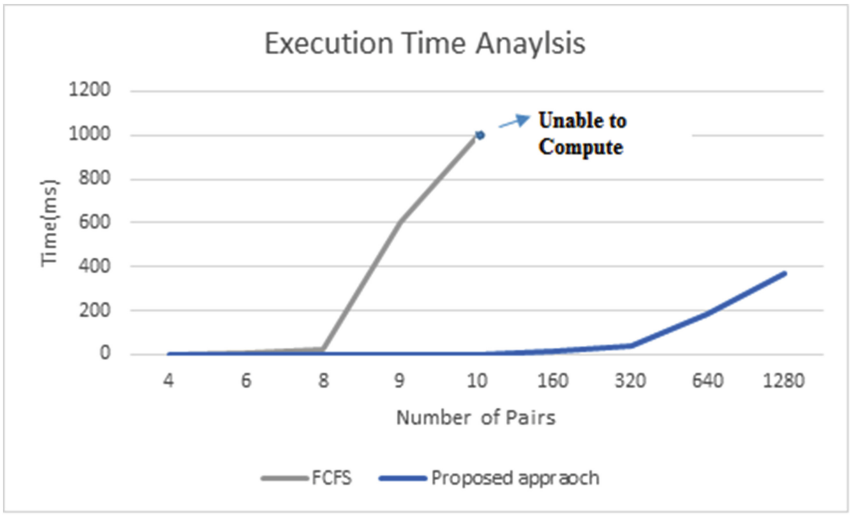


Fig. 5. Comparison of execution time analysis

The proposed approach is devised in such a way that complete utilization of the resources are carried out. For each activity the resources were allocated with complete fairness. Jain et al. [14], proposed to measure the fairness in terms of quantity which is given in the following Eq. 1.

$$f(X) = \left[\sum_{i=1}^{n} x_i\right]^2 /n \times \sum_{i=1}^{n} x_i^2 \tag{1}$$

where $0 \leq f(X) \leq 1$ is fairness measure of resource allocation and $X = (x_1, x_2, ..., x_n)$ implies the allocated resources, n is the number of resources and

activities and x_i is the amount of resource allocated to individuals $i = 1, 2, ..., n$. A large value of $f(X)$ represents fairer resource allocation from the system perspective. The corresponding results is shown in Fig. 6. Also, by deduction we can say greater the value of fairness implies better the stability, that is Fairness is directly proportion to Stability.

Hence, when compared with FCFS and our proposed approach, the allocation of the resources in FCFS is not good for the pairs from 30 onwards. But, whereas in our approach the allocation is stable even in larger pairs which is shown in Fig. 6. Since, FCFS approach couldn't able to perform the allocation under the same environment where our proposed approach is carried out, we couldn't able to compare the fairness allocation for the larger inputs i.e., from 160 pairs to 1280 pairs. Hence, our proposed approach gives better results for larger inputs which is suitable for IoT environment. Because, IoT assumes huge number of devices are going to take part, it reasons out that the proposed approach of resource allocation performs well. Also, it is well suited for the applications of disaster management where the rescue, recovery operations are critical.

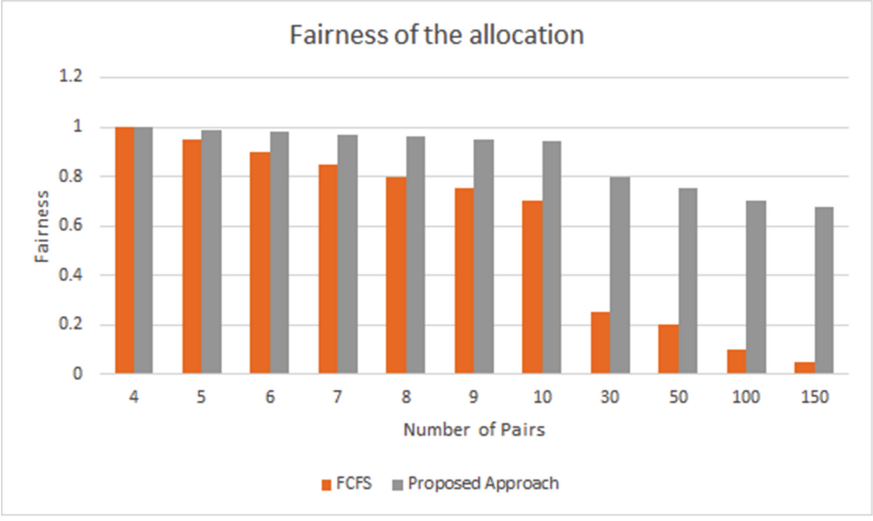


Fig. 6. Comparison of fairness analysis in allocation

Further, without proper mapping to real time environment of identifying the activities and resources, it is hard to validate our approach. Therefore, we made efforts to represent our approach in Google maps [16]. Hence, the proposed approach is shown in real time allocation using google maps as shown in Fig. 7 in which R1..R6 indicates resources and allocated with corresponding activities. The resource r_1 is allocated to activity a_1, resource r_2 is allocated to activity a_3, r_3 to a_5, r_4 to a_2, r_5 to a_6 and r_6 to a_4 respectively. Therefore, our approach assures that all the activities are addressed with proper utilization of all resources.

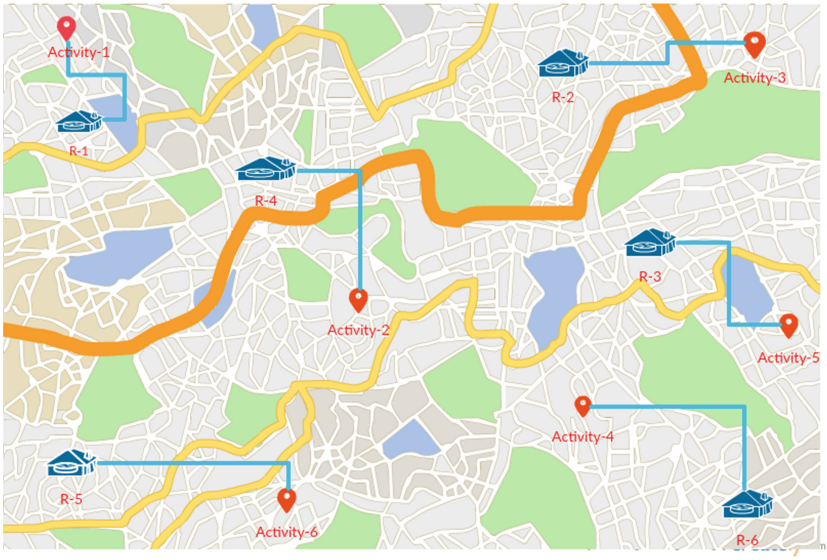


Fig. 7. Real time allocation shown in Google Maps

5 Conclusion

Resource allocation and activity management is critical during disaster scenarios. The use of IoT in the establishment of communication network in such cases helps in efficient mapping of resources to the network entities. Also, knowing the priorities in resources and activities assist to determine the allocation efficiently using stable marriage matching in order to bring the stability in the network. The proposed approach is evaluated in terms of fairness and execution time, which shows better results than FCFS brute force approach.

Acknowledgments. This work is supported by the Department of Electronics and Information Technology (DeiTY), funded by Ministry of Human Resource Development (MHRD), Government of India (Grant No. 13(4)/2016-CC&BT).

References

1. Wahlstrom, M., Guha-Sapir, D.: The Human Cost of Weather-Related Disasters 1995–2015. UNISDR, Geneva (2015). https://www.unisdr.org/2015/climatechange/COP21_WeatherDisastersReport_2015_FINAL.pdf
2. Data Collection Survey for Disaster Prevention in India, Japan, October 2015. http://open_jicareport.jica.go.jp/pdf/12245155.pdf
3. Lee, G.M., Crespi, N., Choi, J.K., Boussard, M.: Internet of Things. In: Bertin, E., Crespi, N., Magedanz, T. (eds.) Evolution of Telecommunication Services. LNCS, vol. 7768, pp. 257–282. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41569-2_13

4. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. J. Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
5. Muaafa, M., Concho, A.L., Ramirez-Marquez, J.: Emergency resource allocation for disaster response: an evolutionary approach (2014)
6. Yang, L., Yang, S.-H., Plotnick, L.: How the Internet of Things technology enhances emergency response operations. Technol. Forecast. Soc. Change **80**(9), 1854–1867 (2013)
7. Kondaveti, R., Ganz, A.: Decision support system for resource allocation in disaster management. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009), pp. 3425–3428. IEEE (2009)
8. Kumar, J.S., Zaveri, M.A.: Clustering for collaborative processing in IoT network. In: Proceedings of the Second International Conference on IoT in Urban Space, pp. 95–97. ACM (2016)
9. Pearce, L.: Disaster management and community planning, and public participation: how to achieve sustainable hazard mitigation. Nat. Hazards **28**, 211–228 (2003)
10. Kumar, J.S., Zaveri, M.A.: Hierarchical clustering for dynamic and heterogeneous Internet of Things. Procedia Comput. Sci. **93**, 276–282 (2016)
11. Pandey, S.K., Zaveri, M.A: Localization for collaborative processing in the Internet of Things framework. In: Proceedings of the Second International Conference on IoT in Urban Space, pp. 108–110. ACM (2016)
12. Kominers, S.D., Sönmez, T.: Matching with slot-specific priorities: theory. Theor. Econ. **11**(2), 683–710 (2016)
13. Manne, F., Naim, M., Halappanavar, M.: On stable marriages and greedy matchings. In: Proceedings of the SIAM Workshop on Combinatorial Scientific Computing, pp. 1–8. ACM (2016)
14. Jain, R., Chiu, D., Hawe, W.: A quantitative measure of fairness and discrimination for resource allocation in shared systems, digital equipment corporation, Technical report DEC-TR-301, vol. 38 (1984)
15. Arora, H., Raghu, T.S., Vinze, A.: Resource allocation for demand surge mitigation during disaster response. Decis. Support Syst. **50**, 304–315 (2010)
16. Svennerberg, G.: Beginning Google Maps API 3. Apress, New York (2010)

Performance Analysis of 32 × 10 Gbps WDM System Based on Hybrid Amplifier at Different Transmission Length and Dispersion

Dipika Pradhan[1(✉)], Abhilash Mandloi[2], and Sajid Shaikh[1]

[1] Electronics and Telecommunication Engineering Department, JSPM, NTC, Pune, India
deepika01513@gmail.com, sajid0077@gmail.com
[2] Electronics Engineering Department, SVNIT, Surat, India
asm@eced.svnit.ac.in

Abstract. A design for a hybrid optical amplifier is presented in this paper. The performance of DWDM system consisting of hybrid amplifier RAMAN + EDFA for NRZ and RZ data format is investigated. It has been observed that RZ format provides highest quality factor 38.5 dB and OSNR 26.6 dB for 32 channels. We further investigated that the hybrid amplifier provides least Bit error rate $8.9e^{-52}$ and $1.02e^{-92}$ for dispersion 2 ps/nm/km and 16 ps/ns/km respectively. It is observed that the Quality factor is for RZ format is 29.5 dB at 8 ps/nm/km and for NRZ format is 24.1 dB at 6 ps/nm/km.

Keywords: Dense Wavelength Division Multiplexing (DWDM)
Single Mode Fiber (SMF) · Non Return to Zero (NRZ) · Return to Zero (RZ)

1 Introduction

To increase the transmission capacity of optical fiber system, a DWDM system is designed. In order to compensate fiber loss for the optical fiber communication system, a multipump Raman amplifier was designed along with EDFA.

Martini et al. [1] have simulated the performance analysis of multipump Raman amplifier with EDFA for WDM system. The gain variation was compensated within the c band. Kelar [2] observed least BER (10^{-40} and 9.0810^{-18}) at 100 km for dispersion 2 ps/nm/km and 4 ps/nm/km respectively.

In this paper we have extended the work reported [2] by using hybrid amplifier. Singth and Kelar [3] investigated RZ provide good quality factor and acceptable bit error rate. A flat gain has been presented in S band (1460–1490 nm).

Liang et al. has examined and compared various types of EDFA-EDFA hybrid amplifiers. The design of H-WDM EDFA was to keep the output power among digital channels at ≤ 0.2 dB while providing output power of ≥ 60 mW and low noise figure of ≤ 4 Db [4]. A 20 channel S band Raman amplifier is analyzed. A novel high gain wide band hybrid amplifier has been reported [5]. Singh et al. optimized the gain flattening filter and reduce the gain ripple across the frequency range from 190 to 197.9 THz [6]. The multiparameter optimization of Raman amplifier has already

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 225–229, 2018.
https://doi.org/10.1007/978-3-319-73712-6_23

reported [6]. The Gain variation of <4.5 dB has been obtained for L band Raman-EDFA hybrid optical amplifier for DWDM system [7]. Singh compared multi terabits DWDM system at different modulation formats such as NRZ, RZ and DPSK, it is found that RZ format is better than all other types of data format [8]. Kelar et al. optimized the hybrid amplifier using different parameters such as Gain and NF. The system achieves 70 km distance at dispersion 16 ps/nm/km [9].

2 Simulation Setup

In this model, 32 channels are transmitted with 100 GHz, channel spacing t 10 Gbps speed in both RZ and NRZ modulation format. Each input signal is amplified by booster. The DWDM system is design with C band ranging from (1530–1554.8 nm) at 100 GHz channel spacing. The experimental setup of EDFA-RAMAN at different transmission distance is shown in Fig. 1.

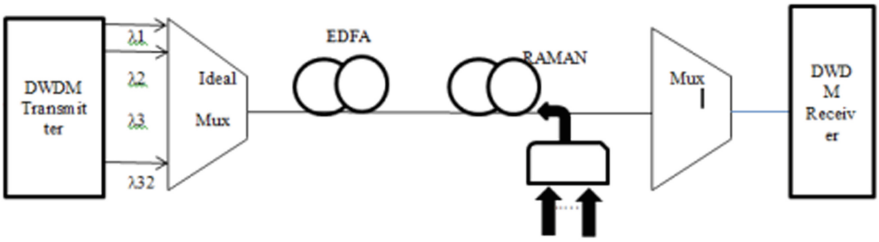


Fig. 1. Simulation setup of Hybrid amplifier in DWDM system

The optical signal is transmitted and measured at different distance for 10 km to 100 km at 2 ps/nm/km to 16 ps/nm/km. Optical power meter and spectrum analyzers are used to measure Q factor and BER. Various parameters are obtained at fixed RAMAN fiber length 20 km, operating temperature is 300 k, pump wavelengths are 211.9 THz, 210.1 THz and 203.5 THz and pump powers are 244.1 mW, 269.9 mW and 60.1 mW respectively (Table 1).

Table 1. General simulation parameters

	Parameters	Value
1	Input signal power	−10 dBm
2	Data rate	10 Gbps
3	Bandwidth	1530–1554.8
4	Band utilized	24.8 nm
5	Modulation format	NRZ and RZ
6	Channel spacing	0.8 nm

3 Result and Discussion

The performance of hybrid amplifier with NRZ and RZ format is compared at different transmission length and dispersion. The system is analyzed at constant input power at −10 dBm. Figure 2 shows that the maximum quality factor is obtained for RZ format is 38.5 Db at 20 km distance of single mode fiber whereas for NRZ is 20.4 Db (Table 2).

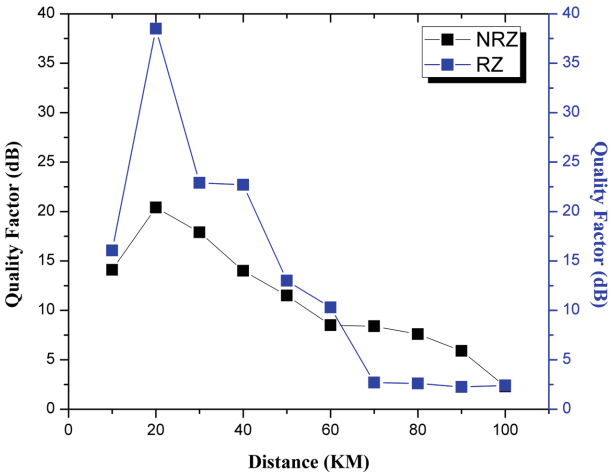


Fig. 2. Quality factor vs Distance of SMF for NRZ format and RZ format

Table 2. Simulation parameters of SMF and EDFA

Sl.no	Parameters	Value (SMF)	Parameters	Value (EDFA)
1	Length	10–100 km	Length	5 m
2	Dispersion	16.75 ps/nm/km	Core radius	2.2 μm
3	Reference wavelength	1555 nm	Er doping radius	2.2 μm
4	Effective area	80 μm^2	Er ion density	1000 ppm-wt
5	Attenuation	0.2 dB/km	Loss at 1550 nm	0.1 dB/m

It is observed that, as the transmission distance increases the quality factor decreases and OSNR also decreases as shown in Fig. 3. From Fig. 4, it is found that the dispersion of single mode fiber is varied from 2 to 16 ps/nm/km and we observed that the Quality factor is for RZ format is 29.5 dB at 8 ps/nm/km and for NRZ is 24.1 dB at 6 ps/nm/km (Table 3).

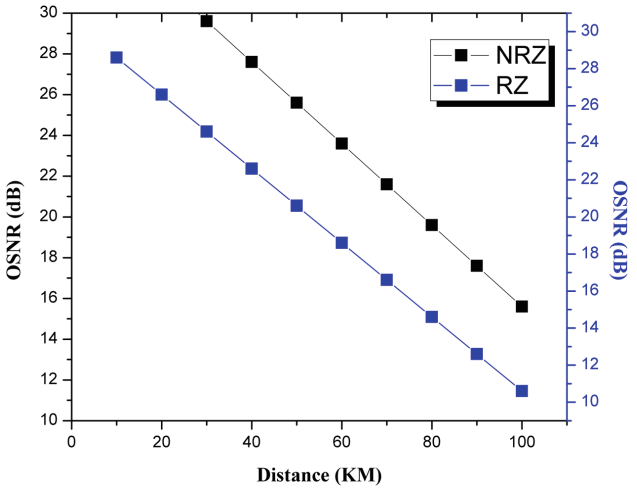


Fig. 3. OSNR vs Distance of SMF for NRZ format and RZ format

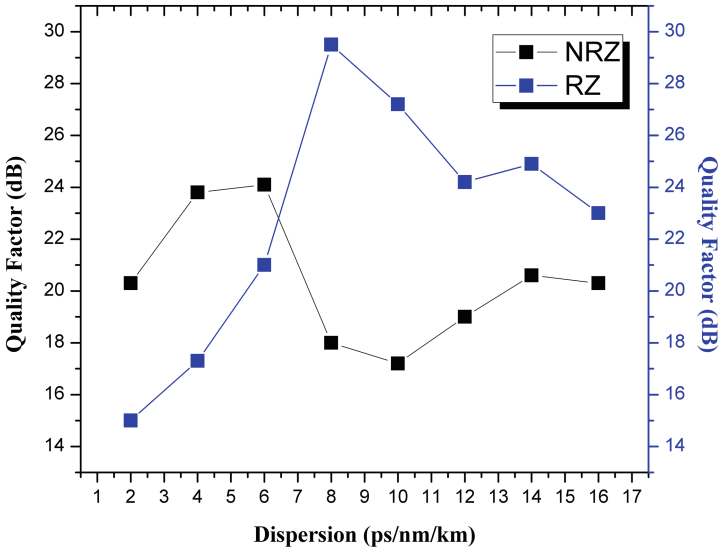


Fig. 4. Quality factor vs Dispersion of SMF for RZ format and NRZ format

Table 3. Simulation parameters of multi pump Raman amplifier.

Pump signal power	Pump signal wavelength	Parameters	Value (Raman amplifier)
		Length	20 km
244.1 mW	1414.5 nm	Dispersion	16.75 ps/nm/km
269.9 mW	1426.5 nm	Effective area	72 μm^2
60.1 mW	1472.5 nm	Attenuation	0.2 dB/km

4 Conclusion

This paper proposed a method for searching least bit error rate and good quality factor for 32 channel 10 Gbps DWDM system. As the transmission distance increases the quality factor and OSNR decreases. The maximum quality factor is obtained 38.5 dB and BER 0 for RZ modulation format at −10 dBm input power. Comparing for both RZ and NRZ data format for hybrid amplifier the RZ format provide better performance than NRZ data format.

References

1. Martini, M.M.J., Castellani, C.E.S., Pontes, M.J.: Gain profile optimization for RAMAN + EDFA hybrid amplifiers with recycled pumps for WDM systems. J. Microw. Optoelectron. Electromagn. Appl. **l9**(2), 100–112 (2010)
2. Kelar, R.S.: Simulation of 16 × 10 GHbps WDM system based on optical amplifiers at different transmission distance and dispersion. Optik **123**, 1654–1658 (2012)
3. Singth, S., Kelar, R.S.: Performance evaluation of 64 × 10 Gbps and 96 × 10 Gbps DWDM system with hybrid optical amplifier for different modulation formats. Optik **123**, 2199–2203 (2012)
4. Liang, T.S., Hsu, S.: The L-band EDFA of high clamped gain and low noise figure implemented using fiber brag grating and DP method. Opt. Commun. **281**, 1134–1139 (2008)
5. Sivanantha Raja A., Vigneshwari, S., Selvendran, S.: Novel high gain and wide band hybrid amplifier designed with a combination of an EYDFA and a discrete Raman amplifier. J. Opt. Technol. **83**(4), 69–79 (2016)
6. Singh, S., Saini, S., Kaur, G.: On the optimization of Raman amplifier using Genetic algorithm in the scenario of a 64 nm 320 channels DWDM system. J. Opt. Soc. Korea **18**(2), 118–123 (2014)
7. Singh, S., Kelar, R.S.: Flat gain L-band Raman-EDFA hybrid amplifier for dense wavelength division multiplexed system. IEEE Photonics Lett. **25**(3), 250–252 (2013)
8. Singh, S., Kelar, R.S.: Performance analysis of 64 × 10 Gbps and 96 × 10 Gbps with DWDM hybrid optical amplifier for different modulation techniques. Optik **123**, 2199–2203 (2012)
9. Kelar, R.S.: Optimization of Hybrid RAMAN/fiber doped fiber for multi terabits WDM system. Optik **124**, 575–578 (2013)

A Review on Poly-Phase Coded Waveforms for MIMO Radar with Increased Orthogonality

Pooja Bhamre(✉) and S. Gupta

ECED, SVNIT, Surat 395007, India
poojamkhairnar@gmail.com, sgupta@eced.svnit.ac.in

Abstract. Multiple input multiple output (MIMO) radar efficiently works with orthogonal signals. Designing of these orthogonal signals affect radar parameters such as range resolution, angle resolution, doppler resolution etc. The paper represents survey of various optimization algorithms used for designing polyphase waveforms. Autocorrelation main lobe width and side lobe peak influence the pulse compression goodness of a code. Autocorrelation side lobe peak(ASP) and cross correlation peak(CP) parameters of code sets are compared.

Keywords: MIMO radar · Optimization · Polyphase waveform

1 Introduction

Compared to standard phased-array radar systems, MIMO radar systems being capable of sending an independent waveforms from various transmit antennas, offer more degrees of freedom which leads to improved angular resolution and parameter identifiability [2], and provides more flexibility for transmit beam pattern design. Parameter identifiability increases M_t times as compared to phased array radar where M_t is number of transmit antennas. The estimation of several target parameters such as range, Doppler, and Direction-of-Arrival (DOA) etc. are the main issue of interest. Since the information of the targets is obtained from the echoes of the transmitted signals, it is straightforward that the design of the waveforms plays an important role in the system accuracy.

The resolution,distance,characteristics of the received signal etc. strongly depends on the shape of the pulse. To achieve better performance in the context of resolution, pulse compression techniques can be employed. Digital pulse compression techniques such as binary phase codes, polyphase codes and frequency codes are judged by their autocorrelation properties [8]. Mutually orthogonal waveforms should be transmitted through various transmit antennas in order to avoid interference as well as acquiring independent information from various target returns.

Our aim is to achieve high range resolution and multiple target resolution. Hence we must concentrate on designing sequences with good autocorrelation and cross-correlation properties. Autocorrelation side lobe peak (ASP) and cross-correlation (CP) must be as small as possible. Smaller ASP contribute towards

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 230–236, 2018.
https://doi.org/10.1007/978-3-319-73712-6_24

reduction in the probability of false alarm, while a narrower main lobe enhances range resolution. Cross correlation property has a significant role in lowering the probability of intercept [3].

The rest of the paper is organized as follows. In Sect. 2 polyphase waveform design problem is formulated. Literature survey of Optimization Algorithms are introduced in Sect. 3. Simulation results are presented in Sect. 4 and finally Sect. 5 concludes the paper.

2 Polyphase Coded Waveform

The designed complex phase waveforms should have a property of constant modulus over all time duration. Each waveform with code length N consists of N samples. The l^{th} waveform of the set of L orthogonal polyphase waveforms is represented by

$$s_l(t) = e^{2\pi j \Phi_l(n)/M} \tag{1}$$

where $\Phi_l(n) \in (0 \leq \Phi_l(n) \leq (M-1))$ with M distinct phases, $l = 1, 2, 3, ..., L$ and $n = 1, 2, 3, ..., N$. L represents the maximum number of radar stations can be accommodated in the radar system. The Autocorrelation function of polyphase sequence s_l with discrete time index k can be represented as

$$\begin{aligned} A(\Phi_l, k) &= \frac{1}{N} \sum_{n=1}^{N-k} e^{j[\Phi_l(n) - \Phi_l(n+k)]} = 0, \quad for \qquad 0 < k < N \\ &= \frac{1}{N} \sum_{n=-k+1}^{N} e^{j[\Phi_l(n) - \Phi_l(n+k)]} = 0, \quad for \qquad -N < k < 0 \end{aligned} \tag{2}$$

Similarly cross correlation function of sequences s_p and s_q is given as

$$\begin{aligned} C(\Phi_p, \Phi_q, k) &= \frac{1}{N} \sum_{n=1}^{N-k} e^{j[\Phi_q(n) - \Phi_p(n+k)]} = 0, \quad for \qquad 0 \leq k < N \\ &= \frac{1}{N} \sum_{n=-k+1}^{N} e^{j[\Phi_q(n) - \Phi_p(n+k)]} = 0, \quad for \qquad -N < k < 0 \end{aligned} \tag{3}$$

The orthogonal waveforms should be chosen which have low autocorrelation side lobe peak and low cross correlation peak. It is very difficult to design three or more polyphase code sets which are having low cross correlation. Different optimization algorithms have been applied previously to not only minimize ASP and CP but also minimize total autocorrelation side lobe energy and cross correlation energy.

The cost function for optimization problem of minimizing ASP and CP given by Deng [4] is as follows:

$$CF_1 = \sum_{l=1}^{L} \max_{k \neq 0} |A(\Phi_l, k)| + \lambda \sum_{p=1}^{L-1} \sum_{q=p+1}^{L} \max_{k} |C(\Phi_p, \Phi_q, k)| \tag{4}$$

Where λ represents weighing factor between autocorrelation and cross correlation function. He carried simulation results by giving equal weight to both the functions. Simulation results resembles that the location of ASP and CP varies with sequences in the optimization process producing abnormal results. For maintaining stability in optimization process, Deng [4] uses cost function considering total energy of autocorrelation side lobes and cross correlation function given by

$$CF_2 = \sum_{l=1}^{L} \sum_{k=1}^{N-1} |A(\Phi_l, k)|^2 + \lambda \sum_{p=1}^{L-1} \sum_{q=p+1}^{L} \sum_{k=-(N-1)}^{N-1} |C(\Phi_p, \Phi_q, k)|^2 \tag{5}$$

He carried out a statistical simulated annealing (SA) algorithm for minimizing CF_2. Results showed that ASP reduces at a rate of $1/\sqrt{N}$ for larger value of $N(> 400)$ allowing more degrees of freedom to minimize cost function. Liu et al. [6] applied Genetic Algorithm (GA)to minimize the following cost function which is side lobe peak and energy based function.

$$\begin{aligned} CF = & \sum_{l=1}^{L} \max_{k \neq 0} |A(\Phi_l, k)| + \sum_{p=1}^{L-1} \sum_{q=p+1}^{L} \max_{k} |C(\Phi_p, \Phi_q, k)| \\ & + \sum_{l=1}^{L} \sum_{k=1}^{N-1} |A(\Phi_l, k)|^2 + \sum_{p=1}^{L-1} \sum_{q=p+1}^{L} \sum_{k=-(N-1)}^{N-1} |C(\Phi_p, \Phi_q, k)|^2 \end{aligned} \tag{6}$$

For the given values of N, L, and M equally weighted objective functions in Eq. (6) can be minimized. The generated polyphase sequences are automatically constrained by Eqs. (1) and (3).

3 Optimization of MIMO RADAR Waveforms

The population-based optimization algorithms to find near-optimal solutions to the difficult multi-objective optimization problems are used in the literature. Population-based algorithms has advantage of employing fewer control parameters. Aubry et al. [1] reported case studies to design waveforms which maximizes the detection probability considering scenarios like signal-dependent or signal-independent interference. Integrated side lobe and peak side lobe of the cross correlation function for different iterations were presented. Mathematical operations required for the calculation of cost function and constraint function are formulated in [9]. The cost function containing integrated side lobe level ratio and peak side lobe level ratio was optimized using genetic algorithm in [7].

As observed from the Table 1, if more weight ($\lambda = 2$) is given to cross correlation energy in the cost function and all other parameters remained same, then reduction in CP value by 15% is observed as expected. But at the same time ASP value has increased by 13.33% [4]. ASP and CP are also inversely proportional to code length N. Slight improved results can be observed using Genetic Algorithm followed by iterative search method [6]. Iterative search continues till

Table 1. Comparison of optimization algorithms

Author	Year	Algorithm	Parameters	Avg. ASP		Avg. CP	
				Normalized value	in dB	Normalized value	in dB
Deng [4]	2004	Integration of SA with traditional iterative code selection	$L = 4, M = 4, N = 40, \lambda = 1$	0.15	−16.5	0.2	−14
			$L = 4, M = 4, N = 40, \lambda = 2$	0.17	−15.39	0.17	−15.39
			$L = 3, M = 4, N = 128, \lambda = 1$	0.0895	−20.96	0.1113	−19.07
Liu et al. [6]	2006	Integration of GA with traditional iterative code selection	$L = 4, M = 4, N = 40, \lambda = 1$	0.147	−16.7	0.2078	−13.64
Reddy and Uttarakumari [10,11]	2013	PSO	$L = 4, M = 4, N = 40, \lambda = 1$	0.1384	−17.1	0.2018	−13.9
	2014	Modified Ant colony algorithm	$L = 4, M = 4, N = 40, \lambda = 1$	0.129	−17.76	0.2068	−13.68
			$L = 4, M = 4, N = 256, \lambda = 1$	0.0039	−48.16	0.00139	−57.14

no phase change is observed. Reddy and Uttarakumari [10] worked out Particle Swarm Optimization (PSO) and got the improved ASP as well as CP, which reduced by 6% and 3% respectively as compared to [6].

Modified ant colony algorithm in which optimized sequence is followed by hamming scan algorithm. It looks for all hamming neighbours of the sequence which further reduces the objective function. By increasing the code length from 40 to 256, reduction in ASP from −17.76 dB to −48.16 dB and CP from −13.68 dB to −57.14 dB is observed. Increasing the code length significantly improves the result but also increases time complexity.

4 Simulation Results

Simulations are carried out in MATLAB by considering the optimized sequences with 4 code sets [11] having orthogonality property. Polyphase waveform with 4 phases $\{0, \pi/2, \pi, 3\pi/2\}$ and code length of 40 is considered. Figure 1 shows Autocorrelation function of every code set, while Fig. 2 describes Autocorrelation function in decibels(dB). ASP and CP are normalized with respect to sequence length.

The pulse compression goodness of a code is decided by its autocorrelation function since in the absence of noise, the output of the matched filter is proportional to the code autocorrelation. The main lobe width (compressed pulse width) and the side lobe levels for the given autocorrelation function of a certain code are the two factors that need to be considered in order to evaluate the codes pulse compression characteristics.

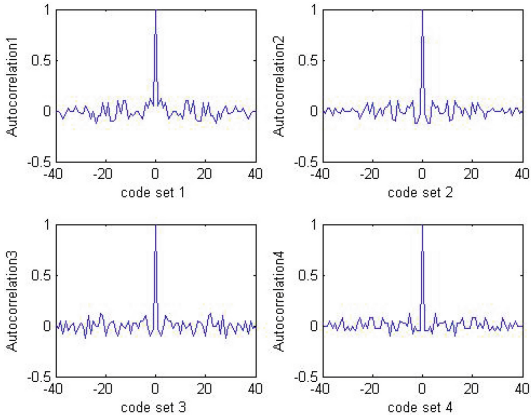


Fig. 1. Autocorrelation function for $L = 4, M = 4, N = 40$.

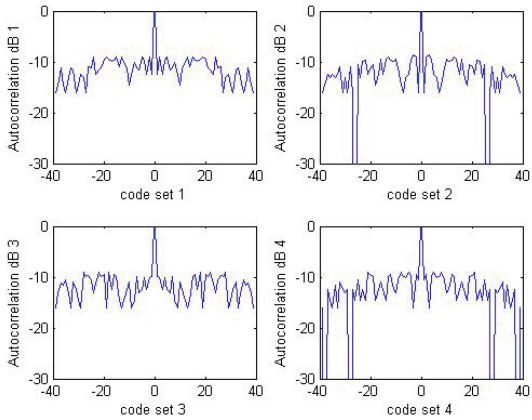


Fig. 2. Autocorrelation function in dB.

The signal design problem for MIMO radar is to generate a pulse with a sharp autocorrelation and low cross correlation. The average ASP from Fig. 1 is 0.129 indicating high probability of detecting weak targets and clutter with the main target. This is because main lobe of the autocorrelation function of weak target or clutter may hide behind side lobe peak of main target which infers miss detection.

The cross correlation function and it's dB equivalents are plotted in Figs. 3 and 4 respectively. The designed waveforms for MIMO radar should have even lower cross correlation in order to detect multiple targets [5]. Cross-correlation peak (CP) gives measure of orthogonality between signals from different antennas. Lesser the CP, lesser the interference between waveforms of different antennas.

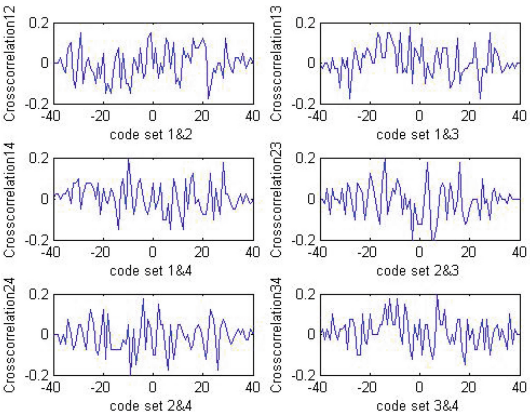


Fig. 3. Crosscorrelation function between code sets.

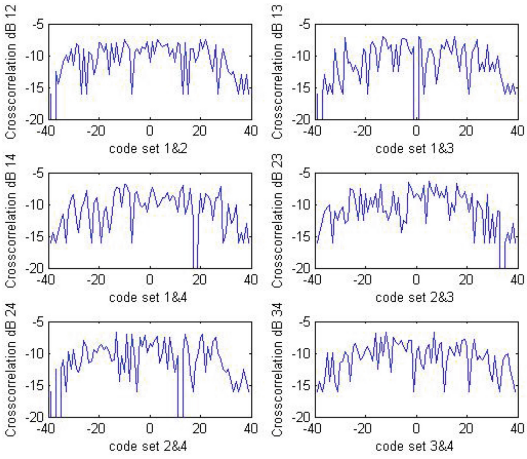


Fig. 4. Crosscorrelation function in dB.

5 Conclusion

The paper compares optimization algorithms (PSO, GA, SA, Modified ant colony) for designing of polyphase waveforms used in MIMO radar. Modified ant colony algorithm [11] appears to be the best among all as far as ASP is concerned. The gradual decrement was observed in ASP and CP values as code length increases. Providing more weights to either ASP or CP in the cost function also improves their values but at the cost of other.

References

1. Aubry, A., Carotenuto, V., Maio, A.D., Farina, A., Pallotta, L.: Optimization theory-based radar waveform design for spectrally dense environments. IEEE Aerosp. Electron. Syst. Mag. **31**(12), 14–25 (2016)
2. Chen, C.Y., Vaidyanathan, P.P.: Compressed sensing in mimo radar. In: 2008 42nd Asilomar Conference on Signals, Systems and Computers, pp. 41–44 (2008)
3. Deng, H.: Synthesis of binary sequences with good autocorrelation and crosscorrelation properties by simulated annealing. IEEE Trans. Aerosp. Electron. Syst. **32**(1), 98–107 (1996)
4. Deng, H.: Polyphase code design for orthogonal netted radar systems. IEEE Trans. Signal Process. **52**(11), 3126–3135 (2004)
5. Gao, C., Teh, K.C., Liu, A., Sun, H.: Piecewise lfm waveform for mimo radar. IEEE Trans. Aerosp. Electron. Syst. **52**(2), 590–602 (2016)
6. Liu, B., He, Z., Zeng, J., Liu, B.: Polyphase orthogonal code design for mimo radar systems. In: 2006 CIE International Conference on Radar, pp. 1–4 (2006)
7. Mehany, W., Jiao, L., Hussien, K.: Polyphase orthogonal waveform optimization for mimo-sar using genetic algorithm (2014)
8. Mow, W.H., Li, S.Y.R.: Aperiodic autocorrelation and crosscorrelation of polyphase sequences. IEEE Trans. Inf. Theory **43**(3), 1000–1007 (1997)
9. Patton, L.K., Frost, S.W., Rigling, B.D.: Efficient design of radar waveforms for optimised detection in coloured noise. IET Radar Sonar Navig. **6**(1), 21–29 (2012)
10. Reddy, B.R., Uttarakumari, M.: Target detection using orthogonal polyphase mimo radar waveform against compound gaussian clutter. Procedia Eng. **64**, 331–340 (2013). International Conference on Design and Manufacturing (IConDM2013)
11. Reddy, B.R., Uttarakumari, M.: Design of orthogonal waveform for mimo radar using modified ant colony optimization algorithm. In: 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2554–2559 (2014)

Designing of SDR Based Malicious Act: IRNSS Jammer

Priyanka L. Lineswala(✉) and Shweta N. Shah

Department of Electronics and Communication, SVNIT, Surat, India
plineswala@gmail.com, snshah@eced.svnit.ac.in

Abstract. Indian Regional Navigation Satellite System (IRNSS) is the regional navigation system designed by India which is identical to well-known Global Position System (GPS). The system promises to provide accurate Position, Velocity and Time (PVT) estimations. In future different applications of Internet on Things (IoT) like smart power distribution grids, sensor networking, vehicular network and airplane navigation systems will be depend on IRNSS. To provide reliable and faithful navigation service, IRNSS is developed by India. But it is highly susceptible to a range of threats like jammer. Here, the different types of jammers are classified in detail based on user, structure and signal. Such jammers are developed by Software Define Radio (SDR) just for experimental purpose. The empirical results are compared with jammer which is available in market.

Keywords: IRNSS · Jammer · Software Define Radio

1 Introduction

Precise location as well as accurate timing information is provided by Global Navigation Satellite System (GNSS). The usage of GNSS is not only for personal car and air craft navigation but, they can be employed for the tracking of birds and animals, to provide automation in different transport agency (like railway, ships) and defense applications. But accuracy and reliability (authority based permission) of such system are very important issues. To solve such issues, Indian Regional Navigational Satellite System (IRNSS) is developed by India. IRNSS from India and Quasi-Zenith Satellite System (QZSS) from Japan is independent and autonomous regional navigation system which provides accurate Position, Velocity and Time (PVT) same as GNSS.

In addition to this, new IRNSS applications are currently under development [1]. For example, IRNSS Satellites are launched and functioning of system is under observation. Some type of applications like "toll collection unit" needs to collect information of IRNSS users, which introduces privacy issues. This motivates the development and use of devices like jammer which can deny IRNSS signal reception.

Jammers are illegal but still in the market different types of jammer are easily available. Analyzing jammer is prerequisite to design detection and mitigation techniques of such intentional interference [2]. It is very smooth to analyze jammer if it is simulated on software. As software based simulations are easy to develop with low cost. Also, the major studies are carried out on software, provide flexibility and controllability.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 237–246, 2018.
https://doi.org/10.1007/978-3-319-73712-6_25

Here, the paper is focused on intentional interference like jammer for IRNSS L5 band [3]. The detail classification of jammer is also included. SDR based different experimental jammers are created using GNU and their signal characteristics are compared with original jammers [4]. From the analysis it emerges that the use of mitigation techniques, significantly improves the performance of satellite receivers even in the presence of strong malicious signals. This study is useful to develop mitigation technique by proper realize characteristics of jammer. Software based jammer provides the flexibility in parameters setting to prove the efficiency of any mitigation technique.

The rest of this paper is organized as follows. The classifications of jammers are mentioned in Sect. 2. The results of different SDR based jammers are described in Sect. 3. Also all of them are analyzed with the help of spectrum analyzer and parameters are compared in Sect. 4. Finally the result performance of the different jamming signal and future scopes are summarized.

2 Classification of Jammer

As mentioned in Fig. 1, the detail classification of jammers are done with different ways like based on user, based on physical structure and based on signal characteristics. Based on user, military jammers are available with larger size and civilian jammers are hand

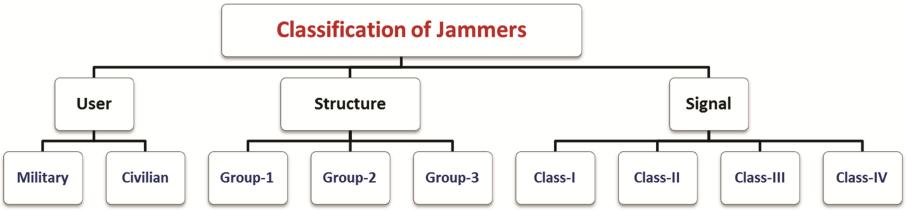


Fig. 1. Classification of jammers

Fig. 2. User based jammer [6]

held device. The Fig. 2 shows the pictorial view of the military based jammers and civilian based jammers.

The requirement of structure depends on the complexity and quality of jammers. Such types of jammers are as shown in Fig. 3. Jammers with auxiliary power supply fall under group 1, jammers with rechargeable battery and external antenna known by group 2. The jammers with rechargeable battery and without external antenna are under group 3 which looks like mobile phone [5].

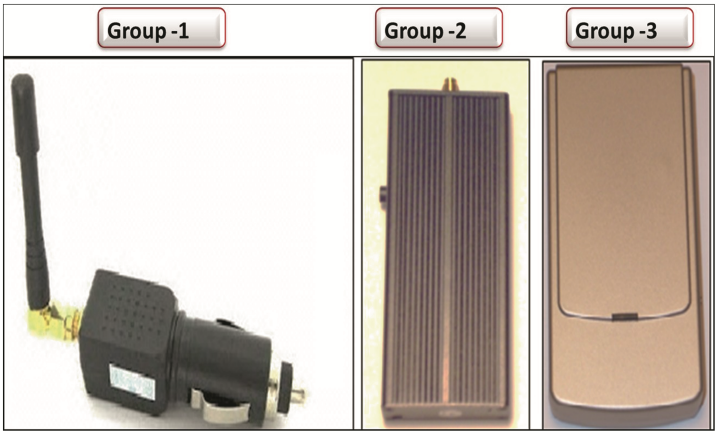


Fig. 3. Structure based jammer [7]

Based on signal, jammers can be classified as (i) Class I: Continuous Wave (ii) Class II: Chirp Signal with Single Saw Tooth (iii) Class III: Chirp Signal with Multiple Saw Tooth (iv) Class IV: Jammers with Frequency Burst. These all types of jammers can be implemented using hardware as well as software [8].

To analyze the jammer it is better to implement these jammers based on software compare to hardware. As software based implementation provides more flexible parameter like power level, frequency value, sweep rate etc. In general as per [9], different signals of jammer can be represented in time and frequency domain as shown in Fig. 5. Figure 4 shows different types of interference signals discussed and implemented here. The left hand side plots show the time domain signals while the right hand side plots refer to frequency domain representation of each signal.

Figure 4(a) illustrates a narrowband CW interference whose frequency is constant within the observation interval. It is a simple CW can be generated easily. Figure 4(b) is a multi-tone interference signal which consists of three frequency components. This type of signal can be generated using multiple signal waveforms with different frequencies. Figure 4(c) is a chirp interference signal whose instantaneous frequency linearly changes over time. Figure 4(d) is a sinusoidal pulse jammer with a 50% duty cycle. The frequency response of this jammer is wider than that of the narrowband CW signal. This type of interference can be generated by simple multiplication of CW with 50% duty cycle pulse or square wave.

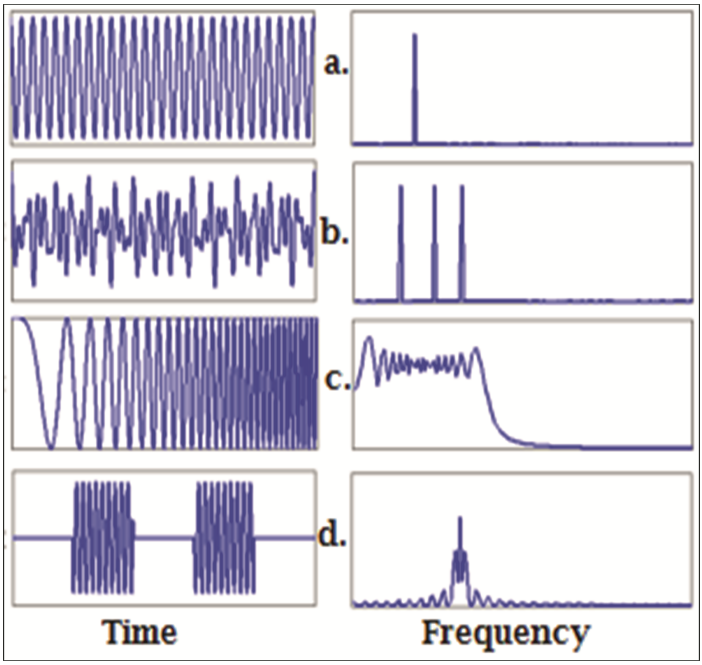


Fig. 4. Characteristics based jammer [9]

3 Implementation of Jammer Using SDR

The different jammers discussed in previous section are implemented by combination of GNU radio software [10] and Amitec SDR hardware [11]. The laboratory set up which was used to implement different class of jammer is as shown in Fig. 5. GNU radio generates jammer signal (laptop) whereas Amitec SDR hardware is transmitting these signals through the antenna. The jammer signal bandwidth and received power were measured by spectrum analyzer CXA N9000A of Agilent Technologies.

Fig. 5. Experiment setup in laboratory

The objective of this study is to use the GNU Radio Companion (GRC) tool to configure the SDR for generating different types of jammers of the IRNSS L5 band and then transmit these signals through the SDR transceiver into or nearer to IRNSS receiver. The SDR connections and settings should be correctly configured for the specified task. More details about SDR implementation is shown in process steps of Fig. 6.

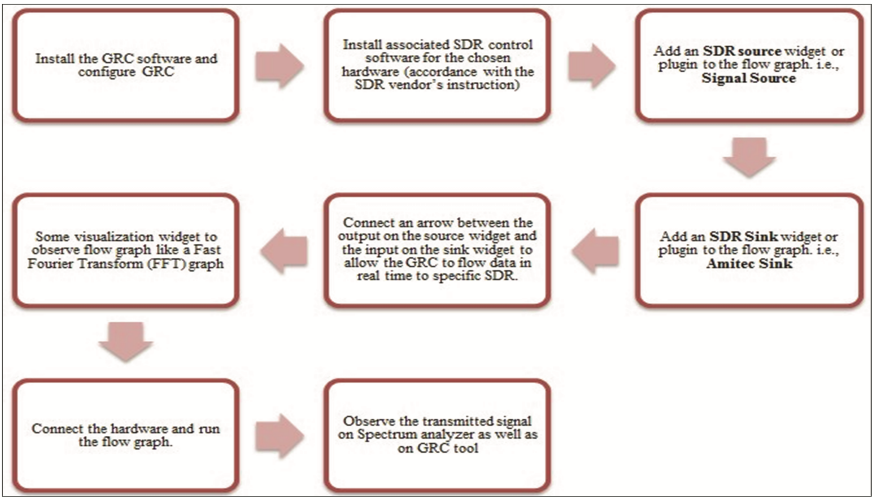


Fig. 6. SDR implementation process flow

3.1 Class I: Continuous Wave Jammer

Simple continuous wave and multi tone jammers are implemented using GNU and SDR which transmit a signal frequency nearer to carrier frequency of IRNSS L5 band. The

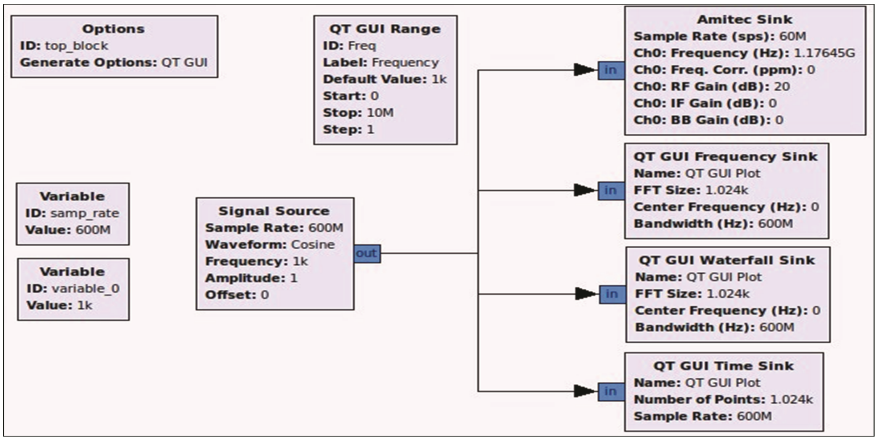


Fig. 7. SDR based single tone continuous wave jammer

flow graph for simulated jammer is shown in Fig. 7. Signal source generating cosine wave and transmit that signal with RF gain 20 dB and frequency 1.17645 GHz. As shown in Fig. 8, the jammer signal bandwidth and received power is measured by signal analyzer from jammer power spectrum. The measured signal bandwidth is around 1 kHz and received power is approximately −34 dBm. The value of bandwidth and power can be changed using GNU flow graph settings.

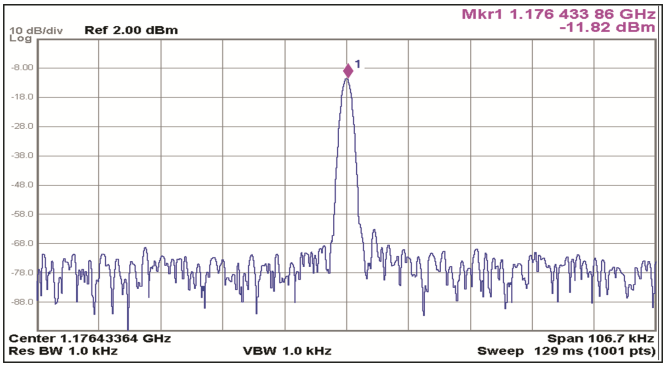


Fig. 8. Single tone CW jammer signal spectrum of IRNSS L5 band

As per [12] Fig. 9 shows same class I narrow band L1 jammer signal spectrum. The frequency of the class I cigarette lighter type PPDs jammer is very close to L1 whereas frequency of generated by SDR jammer is close to L5 based on IRNSS L5 band. The multi tone CW jammer is implemented same as the Fig. 7 but four signal sources are generating cosine wave compare to above single tone CW jammer. From the measured

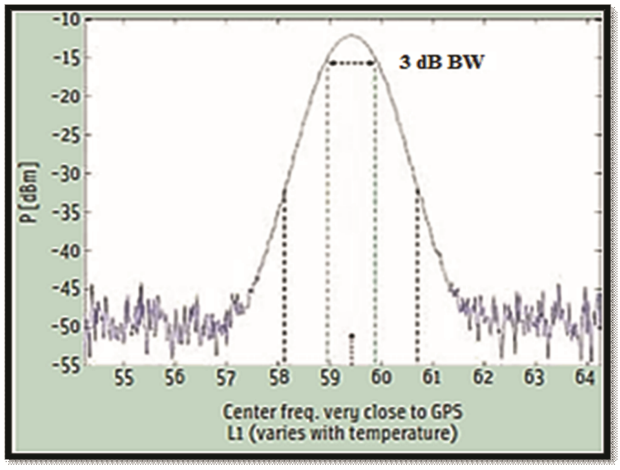


Fig. 9. Single tone CW jammer signal spectrum of GPS L1 band [12]

power spectrum of Fig. 10 power range from −60 dBm to −26 dBm with different peak level of multi tone jammer.

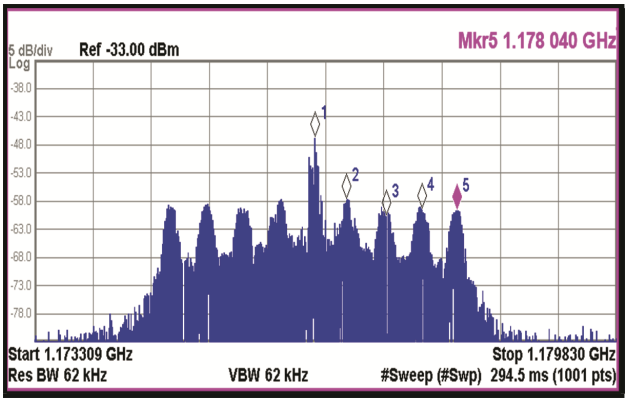


Fig. 10. Multi tone CW jammer signal spectrum of IRNSS L5 band

3.2 Class II: Chirp Signal with Single Saw-Tooth Jammer

This type of jammer signals are made up using Voltage Controlled Oscillator (VCO). Input voltage of VCO varies with a single saw tooth function and the frequency of saw tooth function decides the sweep rate of the resulting signal. However, the upper and lower voltage values decide the bandwidth of the jamming signal. Frequency components of these signals resonate between the higher and lower frequency value with a fix rate of change in frequency [8]. So, chirp signal are much more effective to interfere the navigation signals compare to class I jammer.

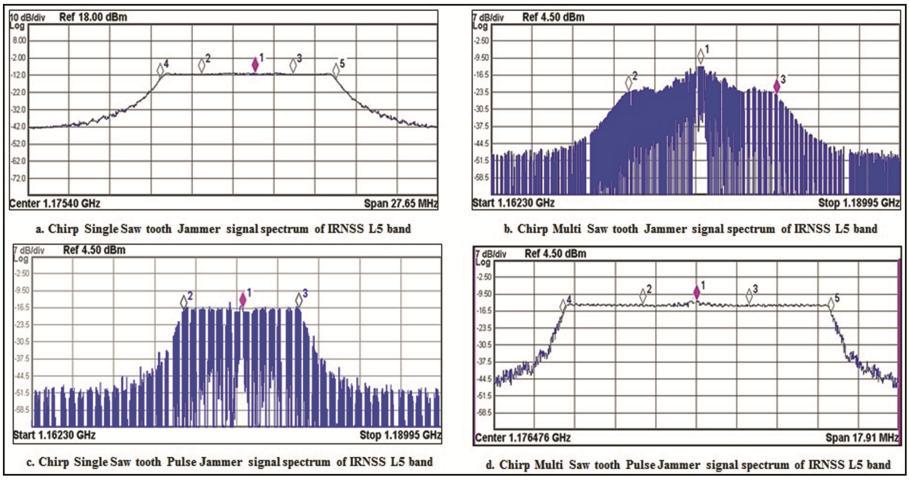


Fig. 11. Power spectrums of Class II, III, IV types of jammer

Here, the saw-tooth signal source is given to VCO and then VCO generates signal frequency of jammer in GNU flow graph. The measured signal bandwidth and power level is shown in power spectrum of Fig. 11(a).

3.3 Class III: Chirp Signal Multi Saw-Tooth Jammer

These signals are same as signal of chirp single saw tooth signals. So, this jammer is implemented same as single saw-tooth jammer using GNU and SDR. However, the multiplication of two saw-tooth signal sources are given to VCO and then VCO generates signal frequency of jammer. In the case of chirp single saw tooth signal it may be possible to detect it and mitigate through a proper technique. Whereas it is very difficult to mitigate multiple saw tooth chirp signals because one cannot predict the newer sweep rate and newer upper and lower value of the frequency [13]. The measured power spectrum of this implemented jammer is shown in Fig. 11(b).

3.4 Class IV: Chirp Signal with Frequency Burst Jammer

These types of jammers are made up of the same signals as it is of Class II or Class III jammers but such generated signals are multiplied with square wave of 10–15 kHz frequency. This multiplication provides burst of frequency and appears like on and off pattern of jammer signal. Also, the chirp signals gets on and off for limited time, in one second more than 10,000 times. These makes nearly impossible to mitigate jammer signals [13].

Square wave for the generation of burst signal is developed in a square wave generator and chirp signals generated independently. The multiplication of both signals is done which implement such jammer as shown in Fig. 12.

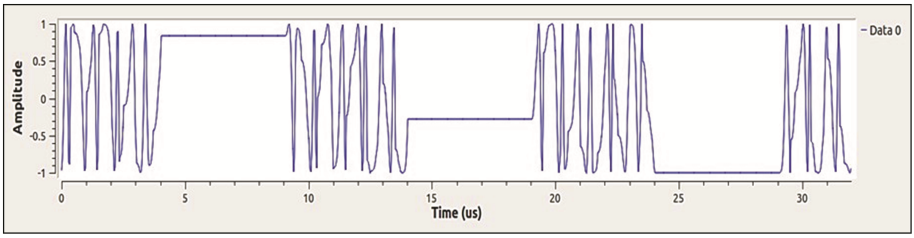


Fig. 12. Implemented chirp signal multi saw-tooth burst jammer

The measured power spectrum shown in Fig. 11(c) and (d) is used to analyze signal bandwidth and power level for the chirp single saw-tooth and chirp multi saw-tooth burst jammer respectively.

4 Analysis of Jammer Parameters

The interference suppression techniques are more beneficial to make reliable and accurate navigation receiver. The examination of these techniques with applications depends on the jammer class. Part of work focuses on implementation of several well-known class of jammer using GNU based SDR. To implement the different jammer support of mathematical equations are considered from reference [9]. In reference [8] jammer parameters are measured values of actual jammer.

So, validation of the work is shown in Table 1 by comparing the parameters of generated SDR jammer signal with reference [8]. The bandwidth values are closer to actual and also the power level controlled by GNU software provides more realistic for consideration.

Table 1. Validation of generated jammer signal parameters with reference [8]

Class	Name	Bandwidth	Power level [8]	Power level
I	Continuous Wave Jammer	~1 kHz	−12.1 dBm	−11.8 dBm
II	Chirp Signal with Single Saw tooth Jammer	~10 MHz	−14.40 dBm	−13 dBm
III	Chirp Signal with Multiple Saw tooth Jammer	~10 MHz	−19.3 dBm	−16.5 dBm
IV	Chirp Signal with Frequency Burst Jammer	~10 MHz	−9.5 dBm	−12.5 dBm

5 Summary

This paper has presented the development of SDR based all type of IRNSS jammers. The importance of navigation signal is based on their application. It is prerequisite to study jammer signal to make the navigation signal receiver robust and more secure against intentional interference like jammer. The brief examples were given to generate jammer based on SDR. Further, such implemented jammer provides parameter wise flexibility for further exploration of intentional interference in laboratory. This work is more useful to make advance research on effective real time mitigation technique for jammer.

Acknowledgements. The author would like to express their thanks to the scientist of SAC, ISRO and Amitec Pvt. Ltd. for their support provided to information about IRNSS and SDR respectively. Special thanks to Mrs. Darshna and Mr. Rutvij for the support provided during measurement campaign.

References

1. Indian Regional Navigation Satellite System: Signal in space ICD for Standard Positioning Services, Version 1.0, ISRO, IRNSS, June 2014
2. Ruparelia, S.M., Lineswala, P.L., Jagiwala, D.D., Desai, M.V., Shah, S.N., Dalal, U.D.: Study of L5 band interferences on IRNSS. In: Proceeding on International GNSS (GAGAN-IRNSS) User Meet, p. 45 (2015)
3. Dovis, F.: GNSS Interference Threats and Countermeasures. Artech House, Norwood (2015)
4. Mitch, R.H., Dougherty, R.C., Psiaki, M.L., Powell, S.P., O'Hanlon, B.W.: Signal characteristics of civil GPS jammers. In: ION GNSS, pp. 1–13 (2011)
5. Grabowski, J.C.: Personal privacy jammers. GPS World **23**, 28–37 (2012)
6. Military Convoy VIP Jammer. http://www.thesignaljammer.com/products/TSJ85W-Vehicle.html
7. Bauernfeind, R., Krmer, I., Beckmann, H., Eissfeller, B., Vierroth, V.: In-car jammer interference detection in automotive GNSS receivers and localization by means of vehicular communication. In: IEEE Forum on Integrated and Sustainable Transportation Systems, 29 June–1 July 2011, Vienna, Austria, pp. 376–381 (2011)
8. Bauernfeind, R., Kraus, T., Sicramaz Ayaz, A., Dtterbck, D., Eissfeller, B.: Analysis, detection and mitigation of InCar GNSS jammer interference in intelligent transport systems, ID: 281260, pp. 1–10. Deutscher Luft- und Raumfahrt kongress (2012)
9. Jahromi, A.J., Broumandan, A., Daneshmand, S., Lachapelle, G.: Vulnerability analysis of civilian L1/E1 GNSS signals against different types of interference. In: ION GNSS, Tampa, FL, pp. 1–10, 14–18 September 2015
10. GNU Radio. http://www.gnuradio.org/redmine/projects/gnuradio.html, http://www.gnuradio.org/redmine/projects/gnuradio/wiki/InstallingGR.html
11. SDR Lab. http://www.sdrlab.com/applications.htmlDd
12. Pullen, S., Gao, G.I.: GNSS jamming in the name of privacy- potential threats to GPS aviation, Inside GNSS Magazine, pp. 34–43, March–April 2012
13. Bauernfeind, R., Eissfeller, B.: Software-defined radio based roadside jammer detector: architecture and results. In: IEEE/ION Position Location and Navigation Symposium (PLANS), California, pp. 1–7, 5–8 May 2014

Sensitivity Analysis of Phase Matched Turning Point Long Period Fiber Gratings

Monika Gambhir(✉) and Shilpi Gupta

Sardar Vallabhbhai National Institute of Technology, Surat 395007, India
gambhirmonika9@gmail.com, shilpigl980@gmail.com

Abstract. This study presents characterization of phase matched turning point long period gratings. It helps in optimizing the grating parameters of these long period gratings viz. grating period, length of grating, for maximum sensitivity. We have calculated spectral variation of refractive indices, effective refractive indices of fundamental and circularly symmetric cladding modes and grating periods. Phase matching curves for first 14 cladding modes have been obtained. Weakly-guiding analysis is used to compute effective refractive indices for the fundamental guided mode and cladding modes. Ultra high sensitivity at turn around points have been verified analytically with the help of general sensitivity factor. LP_{12} cladding mode is observed to be the most sensitive.

Keywords: LPFG (Long Period Fiber Grating)
SRI (Surrounding Refractive Index) · LP (Linearly Polarized)

1 Introduction

In recent years Refractometric sensors based on periodic refractive index modulation in optical fibers namely long period fiber gratings have emerged and illustrated in many applications that include physical parameter sensing [1–3], Adulteration detection [4–6], radiation detection [7], detection of bacteria [8, 9]. These Refractometric sensors offer advantages of being ultra high sensitive, capable for remote sensing and able to operate in harsh environments.

Ultra high sensitivity of long period gratings can be achieved by optimizing the pitch, length and index of modulation of these gratings [10, 11]. Index of modulation is mainly governed by fabrication process. The objective of the article is to compute general sensitivity factor for coupling of fundamental and first 14 cladding modes and finding the parameters at which this factor is maximum. Solutions of the Coupled mode theory based on assumption of weakly guiding regime is used for optimization of long period fiber gratings (LPFGs) for maximum sensitivity.

The paper is organized as follows. In Sect. 2 basic theory of LPFGs is discussed. In Sect. 3 results of the simulations for mode effective refractive indices and their spectral variations have been presented. Grating period calculations at resonant wavelengths have been done to plot phase matching curves. Sensitivity factor has been calculated and its values at turn around points have been highlighted. Finally, conclusions have been drawn in Sect. 4.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 247–253, 2018.
https://doi.org/10.1007/978-3-319-73712-6_26

2 Mode Coupling in LPFGs

The Spatial and periodic modulation of refractive index of the order of hundreds to thousand micrometers in optical fiber causes coupling of fundamental core mode with either co-propagating or counter propagating cladding modes. Propagation characteristics of modes in optical fiber with LPFGs are strong function of the refractive index of surrounding medium.

Coupling of co-propagating fundamental guided mode LP_{01} and cladding modes represented by LP_{0m} takes place in LPFGs according to the phase matching condition, which results in series of attenuated resonance peaks in transmission spectrum [1]. The phase matching condition is given by-

$$\lambda_{res} = \left[n_{effco}(\lambda) - n_{effcl,m}(\lambda)\right]\Lambda \tag{1}$$

where λ_{res} is the resonance wavelength, n_{effco} is effective refractive index of the fundamental mode, $n_{effcl,m}$ id effective refractive index of m^{th} cladding mode and Λ is period of grating.

Phase matching curves drawn between resonance wavelength λ_{res} and grating period Λ indicate presence of turn around point where the slope of the curve changes sign from positive to negative. LPFGs fabricated at these turn around points appear to be ultrahigh sensitive as the slope $\frac{d\lambda_{res}}{d\Lambda}$ of the dispersion curve is infinite. For higher order modes these turn around points occur at higher wavelengths inside the optical communication window.

Mathematical expression for calculating shift in resonance wavelength with respect to change in surrounding refractive index [12] is given by-

$$\frac{d\lambda_{res}}{dn_{surr}} = \lambda_{res}.\gamma.\Gamma_{res} \tag{2}$$

where λ_{res} is resonance wavelength, n_{surr} is surrounding refractive index, γ is general sensitivity factor and Γ_{res} represents surrounding refractive index dependence on waveguide dispersion respectively. Turning points offer ultrahigh sensitivity because a change in wavelength corresponding to change in grating period is infinite.

3 Sensitivity Analysis Results and Discussions

The fiber considered in this paper is of a step-index profile and a three layers structure SMF-28 fiber. The parameters of the fiber are given as: the core radius r1 = 4.61 μm, the cladding radius r2 = 62.5 μm, core region is made up of 3.1% GeO2 doped SiO_2 and cladding region is fused silica. Index of modulation is assumed as 5×10^{-4}.

MATLAB R2008a version has been used as software tool for simulations. Wavelength dependent core and cladding indices have been calculated using Sellmeier equation and the surrounding refractive index is taken as 1.0.

$$n^2(\lambda) = 1 + \sum_{i=1}^{M} \frac{A_i \lambda^2}{\lambda_i^2 - \lambda^2} \tag{3}$$

Table 1. Sellmeier Coefficients for Core and cladding composition of SMF-28 fiber

	SiO_2 (%)	GeO_2 (%)	A_1	λ_1	A_2	λ_2	A_3	λ_3
Cladding	100	0	0.6961663	0.0684043	0.4079426	0.1162414	0.8974994	9.896161
Core	96.9	3.1	0.7028554	0.0727723	0.4146307	0.1143085	0.8974540	9.896161

3.1 Core and Cladding Modes Effective Indices

The effective refractive indices of core and cladding must first be calculated to predict the resonance wavelengths in transmission spectrum of LPFGs. It helps in choosing a particular grating period of LPFG sensor for any application. Interactions between the core and cladding modes at core-cladding boundary and between cladding and surrounding at cladding-surrounding boundary can be solved using Eigen mode/dispersion equation [13].

The propagation is simplified by assuming linearly polarized (LP) modes. Radial power distribution in the core assuming unity power transmitted by the fundamental LP mode is given by-

$$E = \left(\frac{k_0(w)}{J_0(w)}\right) J_0(u) \tag{4}$$

where J_0 and k_0 are Bessel's and modified Bessel functions of first kind, zero order respectively.

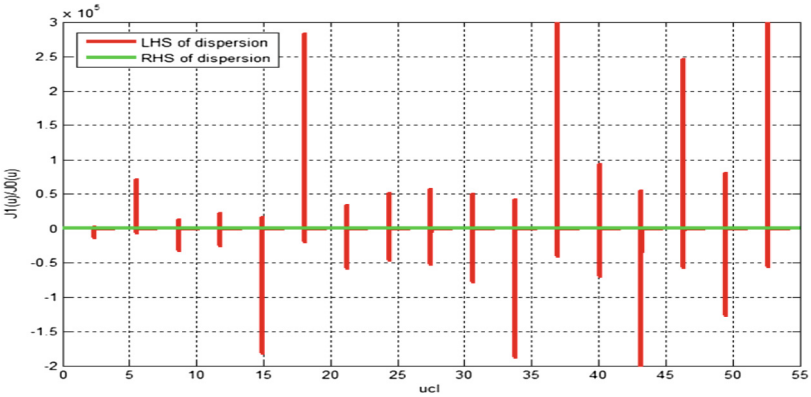


Fig. 1. Cladding mode effective refractive indices at wavelength of 1.3 μm

Eigen value in core u is related to propagation constant β as-

$$u^2 = n_1^2 k^2 - \beta^2 \tag{5}$$

Normalized frequency V is related to waveguide parameters as- $u^2 + w^2 = V^2$. Matching the tangential components of the electric field $\left(E_\varphi, E_z\right)$ and magnetic field $\left(H_\varphi, H_z\right)$ at core cladding interface gives characteristic equation given below which can be solved for Eigen values u & w.

For weakly guided approximation, solution of dispersion equation yields very accurate values of u, w and β [13].

$$u\left(\frac{J_1(u)}{J_0(u)}\right) = w\left(\frac{k_1(w)}{k_0(w)}\right) \tag{6}$$

Evaluation of core and cladding modes effective refractive indices have been accomplished by finding value of waveguide parameter u at which left and right side of dispersion Eq. 5 intersect.

Higher dimensions of cladding results in satisfaction of characteristic equation for a number of u values as indicated in Fig. 1. Using Eq. 5, effective refractive indices of first 14 cladding modes have been calculated. Refractive indices of higher order cladding modes LP_{11} to LP_{14} are given in Table 1.

Table 2. Effective refractive indices of LP_{11} to LP_{14} cladding modes

Wavelength (μm)	n_{effcl_11}	n_{effcl_12}	n_{effcl_13}	n_{effcl_14}
1.1	1.4461076	1.4455052	1.4448490	1.4441390
1.2	1.4444162	1.4436985	1.4429166	1.4420704
1.3	1.4425721	1.4417287	1.4408097	1.4398151
1.4	1.4407757	1.4397962	1.4387289	1.4375731
1.5	1.4388260	1.4377000	1.4364729	1.4351446
1.6	1.4368232	1.4355402	1.4341419	1.4326280
1.7	1.4347670	1.4333164	1.4317353	1.4300233
1.8	1.4325566	1.4309278	1.4291521	1.4272291

3.2 Phase Matching Curves

Sensitivity of LPFGs greatly affected by the choice of grating period. Grating periods at various wavelengths have been calculated. Figure 2(a)–(d) depict phase matching curves for LP_{11}, LP_{12}, LP_{13} and LP_{14} cladding modes and turn around point for these modes appears to occur approximately at 193, 171, 154 and 140 μm respectively.

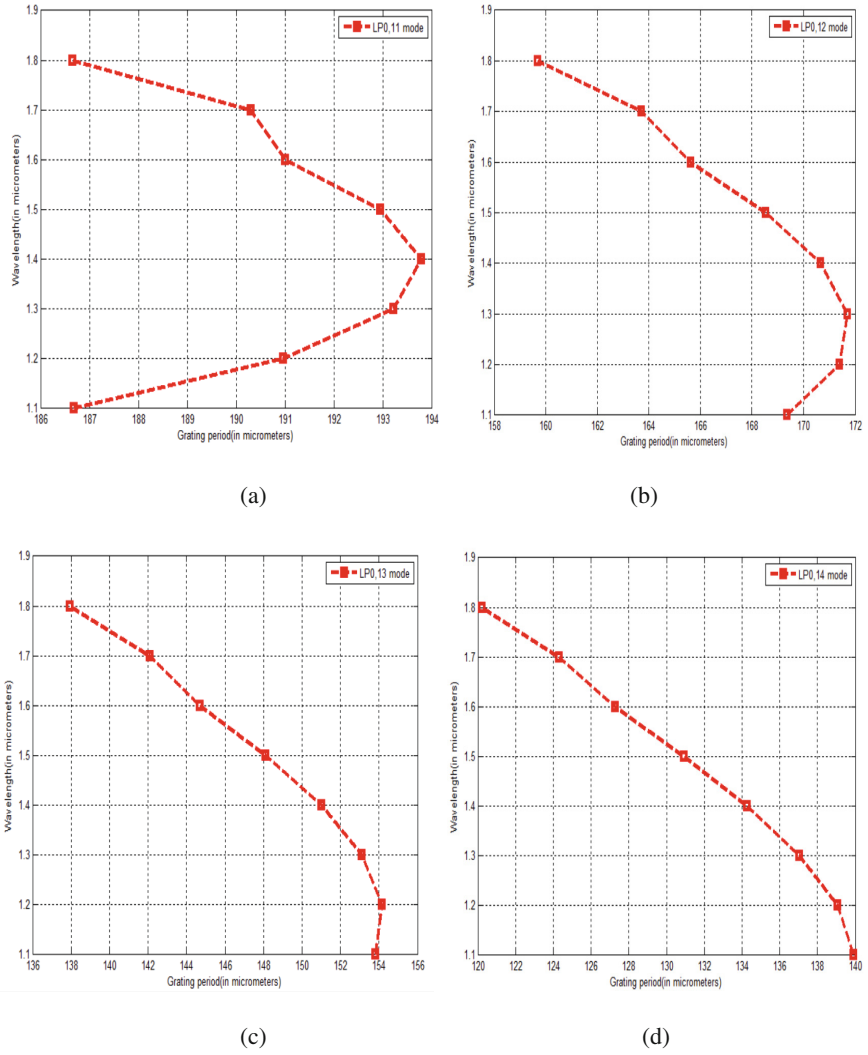


Fig. 2. Phase matching curves of (a) LP_{11} mode (b) LP_{12} mode (c) LP_{13} mode (d) LP_{14} mode

3.3 Sensitivity at TAP in LPFGs

Region of ultrahigh sensitivity for LPFGs can be theoretically demonstrated by magnitude of the sensitivity factor. At turn around points, sensitivity factor exhibits highest values.

In order to compute the sensitivity factor given by Eq. (7) an LPFG, relationship between λ and Λ for m = 1 to 14 cladding modes is calculated.

$$\gamma = \frac{\frac{d\lambda_{res}}{d\Lambda}}{n_{eff}^{co} - n_{eff}^{cl,m}} \tag{7}$$

Highlighted value of sensitivity factor magnitude corresponds to turn around points on phase matching curves for LP_{11} to LP_{14} modes. As highlighted in Table 2, sensitivity factor comes out to be highest at/near turn around points (Table 3).

Table 3. Sensitivity factor for coupling of LP_{11} to LP_{14} cladding modes

Wavelength (in μm)	Sensitivity factor (LP_{11})	Sensitivity factor (LP_{12})	Sensitivity factor (LP_{13})	Sensitivity factor (LP_{14})
1.1–1.2	3.9584	7.6038	**40.1620**	**−14.535**
1.2–1.3	7.0474	**46.2442**	−12.1475	−5.8058
1.3–1.4	**26.3850**	−12.6013	−5.5850	−3.7754
1.4–1.5	−16.4900	−5.7680	−3.7335	−2.8847
1.5–1.6	−6.6045	−3.8680	−2.8805	−2.3747
1.6–1.7	−17.1036	−5.4079	−3.4715	−2.6801
1.7–1.8	−3.0679	−2.3864	−2.0229	−1.7985

4 Conclusion

Characterization of LPFGs require calculations of wavelength dependent refractive indices, mode effective refractive indices and grating periods. An analytical solution has been presented in this paper for calculation of parameters involved in the characterization of gratings in standard SMF-28 fiber. Sensitivity factor is found to be highest for LP_{12} cladding mode. Ultra high sensitivity of long period grating sensors can be fully utilized if parameters are optimized to operate these gratings at turn around points.

References

1. Bhatia, V.: Applications of long-period gratings to single and multi-parameter sensing. Opt. Express **4**, 457–466 (1999)
2. Wang, Y.P., Xiao, L., Wang, D.N., Jin, W.: Highly sensitive long-period fiber-grating strain sensor with low temperature sensitivity. Opt. Lett. **31**, 3414–3416 (2006)
3. Taghipour, A., Rostami, A., Bahrami, M., Baghban, H., Dolatyari, M.: Comparative study between LPFG-and FBG-based bending sensors. Optics Commun. **312**, 99–105 (2014)
4. Kher, S., Chaubey, S., Kishore, J., Oak, S.M.: Detection of fuel adulteration with high sensitivity using turnaround point long period fiber gratings in B/Ge doped fibers. IEEE Sens. J. **13**, 4482–4486 (2013)
5. Mishra, V., Jain, S.C., Singh, N., Poddar, G.C., Kapur, P.: Fuel adulteration detection using long period fiber grating sensor technology. J. Sci. Industrial Res. (JSIR) **46**, 106–110 (2008)

6. Libish, T.M., Linesh, J., Biswas, P., Bandyopadhyay, S., Dasgupta, K., Radhakrishnan, P.: Fiber optic long period grating based sensor for coconut oil adulteration detection. Sens. Transducers **114**, 102–104 (2010)
7. Kher, S., Chaubey, S., Kashyap, R., Oak, S.M.: Turnaround-point long-period fiber gratings (TAP-LPGs) as high-radiation-dose sensors. IEEE Photon. Technol. Lett. **24**, 742–744 (2012)
8. Tripathi, S.M., Bock, W.J., Mikulic, P., Chinnappan, R., Ng, A., Tolba, M., Zourob, M.: Long period grating based biosensor for the detection of Escherichia coli bacteria. Biosens. Bioelectron. **35**, 308–312 (2012)
9. Chiavaioli, F., Biswas, P., Trono, C., Bandyopadhyay, S., Giannetti, A., Tombelli, S., Basumallick, N., Dasgupta, K., Baldini, F.: Towards sensitive label-free immunosensing by means of turn-around point long period fiber gratings. Biosens. Bioelectron. **60**, 305–310 (2014)
10. Śmietana, M., Koba, M., Mikulic, P., Bock, W.J.: Towards refractive index sensitivity of long-period gratings at level of tens of μm per refractive index unit: fiber cladding etching and nano-coating deposition. Opt. Express **24**, 11897–11904 (2016)
11. Esposito, F., Ranjan, R., Campopiano, S., Iadicicco, A.: Experimental study of the refractive index sensitivity in arc-induced long period gratings. IEEE Photonics J. **9**, 1–10 (2017)
12. Shu, X., Zhang, L., Bennion, I.: Sensitivity characteristics of long-period fiber gratings. J. Lightwave Technol. **20**, 255–266 (2002)
13. Oh, K., Paek, U.C.: Silica Optical Fiber Technology for Devices and Components: Design, Fabrication, and International Standards, vol. 240. Wiley, Hoboken (2012)

Performance Analysis of Nakagami and Rayleigh Fading for 2 × 2 and 4 × 4 MIMO Channel with Spatial Multiplexing

Mitesh S. Solanki(✉) and Shilpi Gupta

Electronics Engineering Department, SVNIT, Surat 395007, India
solankimitesh89@gmail.com, shilpig1980@gmail.com

Abstract. This correspondence presents the spectral efficiency analysis of 2 × 2 and 4 × 4 multiple input multiple output (MIMO) system. Analysis in this article has been done over Nakagami-m and Rayleigh fading channel. In this work analytical model as well as simulation has been observed for spectral efficiency of MIMO system over fading environment with an aid of Singular Value Decomposition (SVD) and waterfilling algorithm. It has also been observed the performance of Nakagami-m channel probabilistic model which one-to-one aligning between Nakagami-m and Rayleigh fading distributions in MIMO system.

Keywords: MIMO · Spectral efficiency · Singular value decomposition
Waterfilling algorithm

1 Introduction

The spectral efficiency is elemental parameter in the design of wireless mobile communication systems as it increases the data throughput of the system. Still it is tolerated to fading, which deteriorate the performance of wireless mobile communication system [1, 2]. The normalized capacity of the bandlimited Additive White Gaussian Noise (AWGN) channel, which is due to Shannon limit. However, the cost of reaching the larger spectral efficiency is an increase in the SNR per bit. Therefore, any digital modulation schemes are convenient for communication channels that are bandwidth limited, where is desired a channel capacity to bandwidth ratio greater than one [3]. Power and bandwidth are major challenging resources [4]. The recent perception of multiple-input multiple-output system especially in wireless communication is one of the most impressive interest in research work. So now widely referred to as MIMO technology, this concept can extremely improve data throughput and link performance in wireless networks [5].

The basic facts which makes reliable wireless transmission is time-varying multipath fading [6]. Improving the quality or decreasing the efficient probability of error in a multipath fading channel is very difficult. The large channel capacities associated with MIMO channels are based on the premise that a rich scattering environment, provides independent transmission paths from each transmit to each receive antenna [4]. Now considering linear regression as singular value decomposition provides a MIMO system transmitter and receiver antenna pair to increase the data throughput without increasing

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 254–263, 2018.
https://doi.org/10.1007/978-3-319-73712-6_27

bandwidth usage or transmit power. MIMO system increases throughput linearly with the help of spatial multiplexing [7].

In classical approach of MIMO system, multiple links provides a spatial diversity which leads to linear scale large channel capacity and improved quality of received signal compared to single antenna system [6]. MIMO system can be modelled with a combination of transmit and receive diversity scheme. Three most basic performance evaluations of wireless communication are data throughput, coverage and seamless connectivity.

Under appropriate flat fading environment, having together multiple antenna configurations, provide an added spatial dimension for communication system and yields a degree of- freedom. The channel capacity of such a MIMO channel with n_t transmit and n_r receive antenna is proportional to n, where $n = min(n_t, n_r)$ [6]. In this paper, simulations have been carried out for 2×2 and 4×4 MIMO system for high rich scattering environment using Nakagami-m probabilistic channel model with the help of MATLAB tool.

The flow of this paper is as follows: The wireless system model is defined in Sect. 2. In Sect. 3 exact expression of channel capacity is derived with the help of singular value decomposition and waterfilling algorithm. Section 4 describes fading distributions. Finally, Sect. 5 incorporate the concluding remarks of this proposed research work.

2 System Model

Consider an $n_r \times n_t$ point-to-point MIMO system $n_r = n_t$, where n_t and n_r denote the number of transmit and receive antennas, respectively. Assume channel state information (CSI) to be known perfectly at both ends. Let $\bar{x} = (x_1, \ldots, x_{n_t})$ be the vector of symbols transmitted by the n_t transmit antennas (Fig. 1).

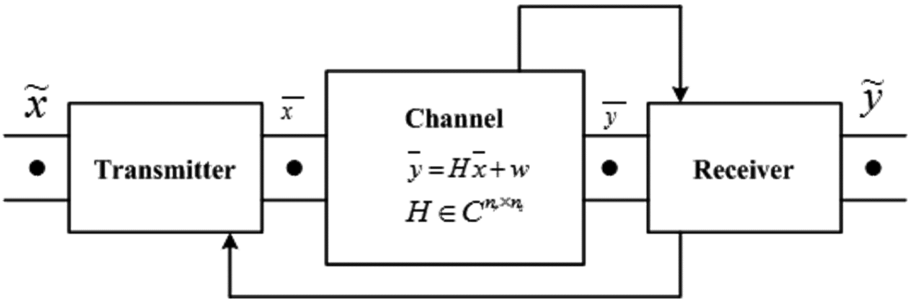


Fig. 1. System model

Let $H = \{h_{ij}\}$, $i = 1, \ldots, n_r$, $j = 1, \ldots, n_t$ be the $n_r \times n_t$ channel matrix are independent and identically distributed (i.i.d) Gaussian variables with zero mean, where h_{ij} is the complex gain between the j^{th} transmit antenna and the i^{th} receive antenna. The $n_r \times 1$ received vector is given by

$$\bar{y} = \mathrm{H}\,\bar{x} + \mathrm{w} \tag{1}$$

where $\bar{x} \in \mathbb{C}^{n_t}, \bar{y} \in \mathbb{C}^{n_r}$ and $w \sim \mathbb{C}\mathcal{N}\left(0, N_0\right)$ denote the transmitter vector, received vector and additive white Gaussian noise vector respectively at a symbol time. The channel matrix $H \in \mathbb{C}^{n_r \times n_t}$ is deterministic. Specifically, assumed the instantaneous channel gains, known as the channel state information, are known perfectly at transmitter or at receiver in order to improve the spectral efficiency and to reduce probability of error [7].

The noise vector is statistically i.i.d complex-valued Gaussian random variables with zero mean. Consider total power constraint $\mathbb{E}\left[\|\bar{\mathrm{x}}\|^2\right] = \mathrm{p}$ at the transmit antennas.

3 MIMO System Channel Capacity

The capacity of MIMO system can be derived with the help of linear regression method which linear transforms the MIMO channel into parallel sub channels. From basic theory of linear algebra, all linear transformation can be derived as a composition of three steps: a rotation, a scaling, and once again rotation. Now consider singular value decomposition of a channel matrix H. A matrix H can be represented as:

$$\mathrm{H} = \mathrm{U}\,\mathrm{D}\,\mathrm{V}^{\mathrm{H}} \tag{2}$$

where $(.)^H$ denotes the conjugate transpose; $V \in \mathbb{C}^{n_t \times n_t}$ and $U \in \mathbb{C}^{n_r \times n_r}$ are (rotation) unitary matrices with left and right singular vectors of H as their columns and $\mathrm{D} \in \mathfrak{R}^{n_r X n_t}$ is a matrix whose diagonal elements are positive real numbers and whose other channel coefficients are zero. The diagonal coefficients $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_{min}}$ are the ordered singular values of the channel matrix H, where $n_{min} := \min\left(n_r, n_t\right)$ and $n_{max} := \max\left(n_r, n_t\right)$. These are important conditions of the singular value decomposition. The number of nonzero singular values is equal to the rank of the channel matrix H [7]. Since the squared singular values λ_i^2 are the eigenvalues of

$$\begin{cases} \mathrm{HH}^{\mathrm{H}}, & \text{if } \mathrm{n_r} \leq \mathrm{n_t} \\ \mathrm{H}^{\mathrm{H}}\mathrm{H}\,, & \text{if } \mathrm{n_r} > \mathrm{n_t} \end{cases} \tag{3}$$

Here, singular values transforms the channel into n_{min} parallel sub channels λ_i, $1 \leq n \leq n_{min}$. Now, it can be expressed as

$$\mathrm{H} = \sum_{\mathrm{i}=1}^{\mathrm{n}\min} \lambda_{\mathrm{i}} \mathrm{u}_{\mathrm{i}} \mathrm{v}_{\mathrm{i}}^{*} \tag{4}$$

It can be seen that the rank of H is precisely the number of diagonal values. It can be defined as

$$\widetilde{y} = U^H \bar{y},\ \widetilde{x} := V^H \bar{x}, \widetilde{w} := U^H w \tag{5}$$

Substituting (2) and (5) into (1) the received signal vector can be rewritten as

$$\widetilde{y} = \mathrm{D}\widetilde{x} + \widetilde{w} \tag{6}$$

where D is diagonal coefficients vector. Thus, the powers of $\bar{x}$ and $\widetilde{x}$ are the same, as well as $\bar{y}$ and $\widetilde{y}$, w and $\widetilde{w}$. The equivalent model of the system can be depicted in Fig. 2, which shows that the MIMO channel is converted into n_{min} parallel sub-channels through SVD.

$$\widetilde{y}_i = \lambda_i \widetilde{x}_i + \widetilde{w}_i \qquad i = 1, 2, \ldots, n_{min} \tag{7}$$

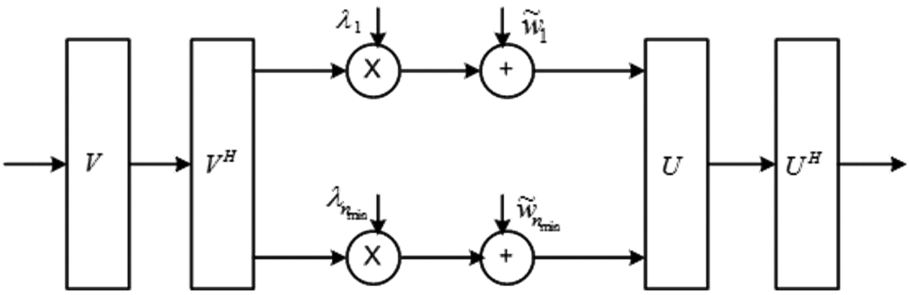


Fig. 2. MIMO system and equivalent model [5]

SVD is obtainable when H is known to both the ends. The optimal power denotes by P_i the power associated with the i^{th} symbol. That is

$$P_i \triangleq \mathbb{E}\left\{ \left|\widetilde{x}\right|^2 \right\} \tag{8}$$

Main goal is to find the optimum distribution for the set $\{P_i\}$ under the constraint that their sum is a fixed value.

$$\sum_{i=1}^{n_{min}} P_i = P \tag{9}$$

Finding the optimum set of $\{P_i\}$ is a standard optimization problem, which can be solved using the Lagrange multiplier technique. The resulting algorithm that implements this solution is called the waterfilling algorithm.

The optimal power allocation strategy has been shown in Fig. 3. For a faithful communication, power allocated to the i^{th} sub channel should be such that the total power constraint is met.

$$\max_{P_1,\ldots,P_{n_{\min}}} \sum_{i=1}^{n_{min}} log\left(1 + \frac{\mathrm{P_i}\lambda_i^2}{\mathrm{N_0}}\right) \tag{10}$$

subject to Eq. (9). The power P is the average power constraint.

$$P_i^* = \left(\frac{1}{\beta} - \frac{N_0}{\lambda_i^2}\right)^+ \tag{11}$$

where β, a constant is called the Lagrange multiplier and satisfies condition

$$\sum_{i=1}^{n_{min}} \left(\frac{1}{\beta} - \frac{N_0}{\lambda_i^2}\right)^+ = P, \tag{12}$$

$$P_i^* = \left(\mu - \frac{N_0}{\lambda_i^2}\right)^+ \tag{13}$$

where, $\mu \triangleq \frac{1}{\beta}$. These Values of P_i^* and μ are used to iteratively compute the set of values for $\{P_i\}$ that maximize the channel capacity. Optimal capacity is achieved by summing the capacities of individual sub channels [7]. Thus,

$$C = \sum_{i=1}^{n_{min}} log_2\left(1 + \frac{P_i^* \lambda_i^2}{N_0}\right) \text{ bps /Hz} \tag{14}$$

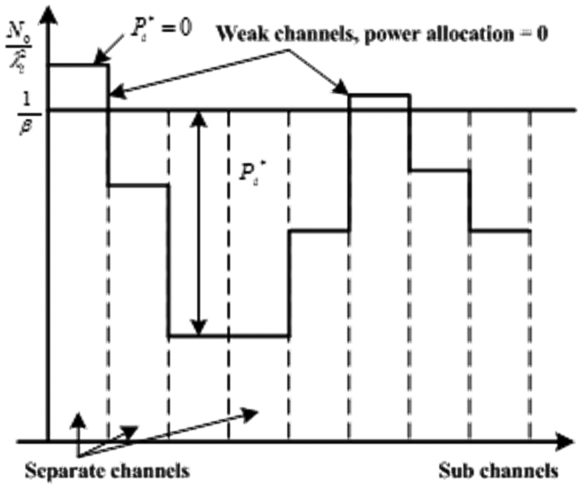


Fig. 3. Waterfilling power allocation [5]

In this way, the MIMO system can explore the spatial multiplexing of multiple streams.

4 Fading Wireless Channels

In this paper consider two fading distributions, Rayleigh and Nakagami-m multipath fading has been considered for analysis.

Rayleigh fading channel: The Rayleigh random variable r has the probability distribution:

$$f_R(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)$$

Where, $\mathbb{E}\{r^2\} = 2\sigma^2$ *and* $r \geq 0$ is the power is exponentially distributed.

Nakagami fading channel: The instantaneous power distribution is expressed as

$$f_R(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m}\exp\left(-\frac{m}{\Omega}r^2\right), r \geq 0$$

where $\{r^2\} = \Omega, m = \frac{\mathbb{E}^2\{r^2\}}{Var\{r^2\}}, \Gamma(.)$ is the Gamma function and Ω denotes average fading power and controls the spread of the distribution.

A close observation of Table 1 reveals that Nakagami-m fading model results in the special case $m = 1$ which towards Rayleigh distribution. For $m > 1$, the fluctuations of the signal reduce compare to Rayleigh fading, and Nakagami-m tends to be more line of sight components.

Table 1. Nakagami and Rayleigh fading distributions for variation of m & σ^2

σ^2	m	Pdf of Nakagami channel	Pdf of Rayleigh channel
0.5	0.5	$8x^3\exp(-2x^2)$	$2x\exp(-x^2)$
	1	$2x\exp(-x^2)$	$2x\exp(-x^2)$
	2	$x\exp(-x^2/2)$	$2x\exp(-x^2)$

5 Simulation Results

Simulation parameters has been shown in Table 2.

Table 2. Simulation parameters

Sr. no.	Parameters	Description
1	Number of iterations	10,000
2	Execution time	9 min
3	Fading channels	Nakagami & Rayleigh
4	SNR range	−10 to 20 dB
5	m values	0.5, 1 & 2

Spectral efficiency for 2 × 2 and 4 × 4 MIMO systems over Nakagami and Rayleigh fading have been shown in Figs. 4 and 5 for m = 0.5, 1 and 2.

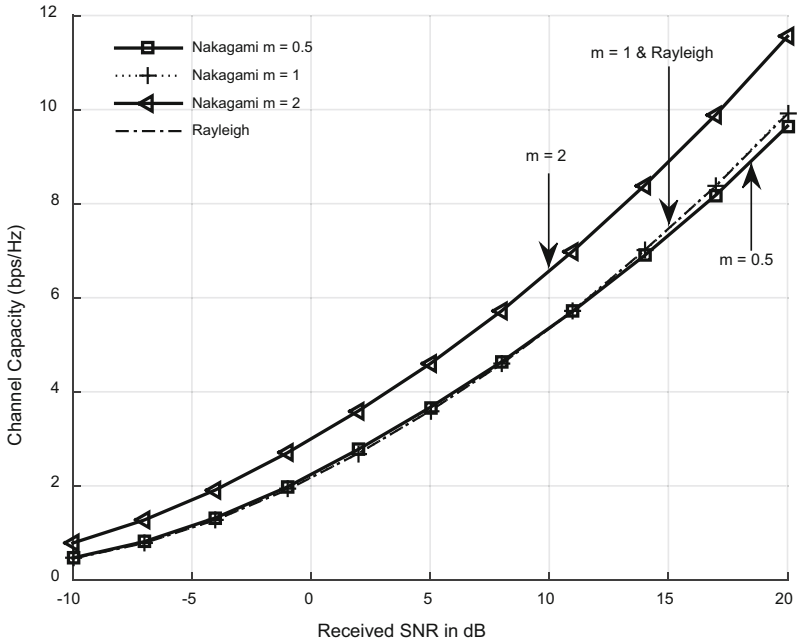


Fig. 4. Spectral efficiency for 2 × 2 MIMO system over Nakagami and Rayleigh fading channel

For 2 × 2 MIMO system it has been observed that, as the received average signal to noise ratio varies from 5 to 20 dB, channel capacity rises from 5 to 11 bps/Hz for Nakagami fading channel and 3 to 9 bps/Hz for Rayleigh fading channel. Similar analysis of 4 × 4 MIMO system reveals that here also channel capacity rises from 7 to 21 bps/Hz in Nakagami fading channel and 6 to 18 bps/Hz in Rayleigh fading channel.

The empirical distributions (histogram) of singular values for 2 × 2 and 4 × 4 MIMO system with 10,000 iteration for random realization of H have been shown in Figs. 6 and 7 respectively.

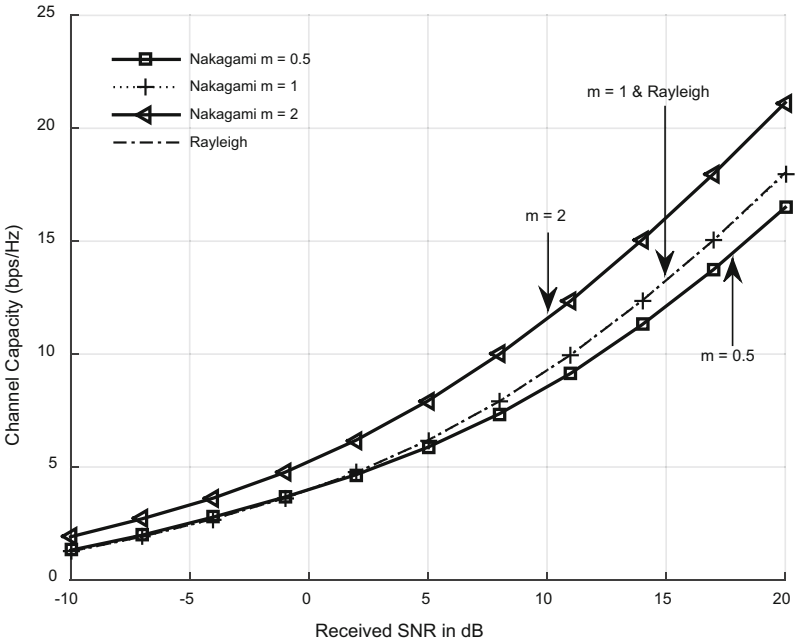


Fig. 5. Spectral efficiency for 4×4 MIMO system over Nakagami and Rayleigh fading channel

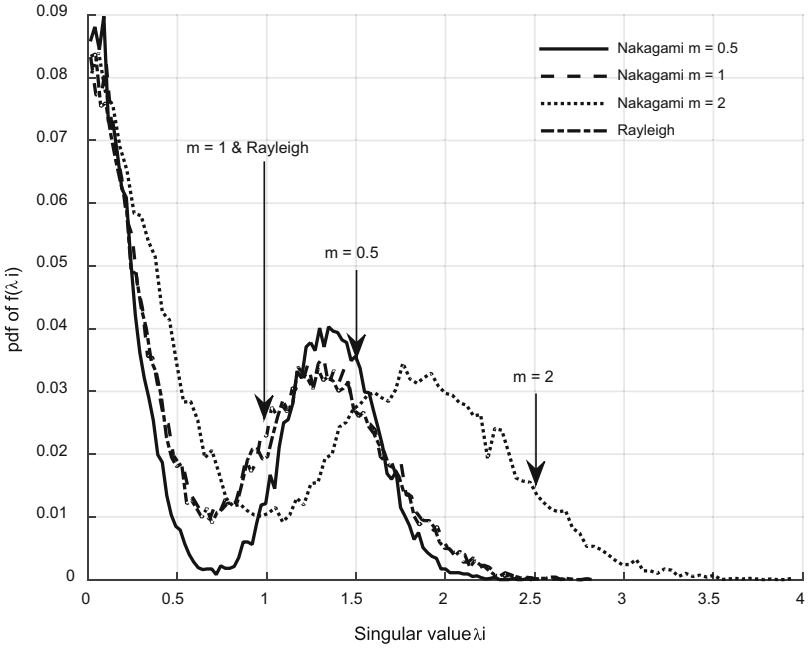


Fig. 6. Probability density function for 2×2 MIMO system over Nakagami and Rayleigh fading

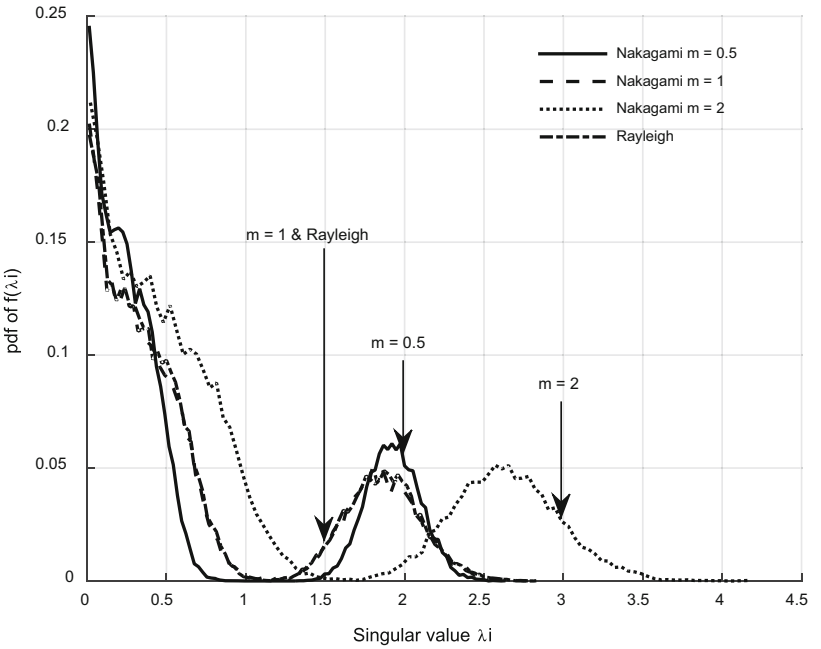


Fig. 7. Probability density function for 4×4 MIMO system over Nakagami and Rayleigh fading

6 Conclusion

In this research work spectral efficiency analysis of 2×2 and 4×4 MIMO systems have been done in both ways analytically as well as through simulation. Here, improvements in MIMO channel capacity and distribution function has been also analyzed among that the variation of parameter m from 0.5 to 2. The Nakagami distribution approaches towards Rayleigh distribution by varying m from 0.5 to 1. From above observations it can be concluded that m parameter is a well suited quantity for characterization of scattering effects.

References

1. Hasan, M.I., Kumar, S.: Spectral efficiency of dual diversity selection combining schemes under correlated Nakagami-0.5 fading with unequal average received SNR. Telecommun. Syst. **64**, 3–16 (2016)
2. Dholakia, P.M., Kumar, S., Vithalani, C.H.: Performance analysis of 4×4 and 8×8 MIMO system, to achieve higher spectral efficiency in Rayleigh & Rician fading distributions. Wirel. Pers. Commun. Int. J. **79**, 687–701 (2014)
3. Proakis, J.G.: Digital Communications, 4th edn. Mc Graw-Hill, New York (2001)
4. Goldsmith, A., Jafar, A., Jindal, N., Vishwanath, S.: Capacity limits of MIMO channels. IEEE J. Sel. Areas Commun. **21**(5), 684–702 (2003)

5. Heath, R.W., Larsson, E.G., Murch, R., Nehorai, A., Uysal, M.: Special issue: Multiple-Input Multiple-Output (MIMO) Communications. Wirel. Commun. Mob. Comput **4**, 693–696 (2004)
6. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. IEEE J. Sel. Areas Commun. **16**(5), 1451–1458 (1998)
7. Tse, D., Viswannath, P.: Fundamental of Wireless Communication. Cambridge University Press, Cambridge (2005)
8. Paulraj, A.J., Gore, D.A., Nabar, R.U., Bolcskei, H.: An overview of MIMO communications —a key to gigabit wireless. Proc. IEEE **92**, 198–218 (2004)
9. Rouffet, D., Kerboeuf, S., Cai, L., Capdevielle, V.: A technical paper on, "4G MOBILE". Alcatel Telecommunication Review, pp. 1–7 (2005)
10. Prasad, R.: Universal Wireless Personal Communication. Artech House, Norwood (1998)
11. Tranter, W.H., Shanmugan, K.S., Rappaport, T.S., Kosbar, K.L.: Principles of Communication System Simulation with Wireless Applications. Prentice Hall Professional Technical Reference, Upper Saddle River (2003)
12. Beaulieu, N.C., Cheng, C.: Efficient Nakagami-m fading channel simulation. IEEE Trans. Veh. Technol. **54**(2), 413–424 (2005)
13. Fraidenraich, G., Leveque, O.: On the MIMO channel capacity for the Nakagami-m channel. IEEE Trans. Inf. Theory **54**, 3752–3757 (2008)
14. Kumar, S., Pandey, A.: Random matrix model for Nakagami-Hoyt Fading. IEEE Trans. Inf. Theory **56**, 2360–2372 (2010)
15. Basnayaka, D.A., Di Renzo, M., Haas, H.: Massive but few active MIMO. IEEE Trans. Veh. Technol. **65**, 6861–6877 (2016)

Wavelet Based Feature Level Fusion Approach for Multi-biometric Cryptosystem

Patel Heena, Paunwala Chirag(✉), and Vora Aarohi

Electronics and Communication Engineering Department, SCET, Surat, India
hpatell323@gmail.com, cpaunwala@gmail.com,
vaarohi@gmail.com

Abstract. Biometric cryptosystems incorporates the benefits of both cryptography as well as biometrics i.e. higher security levels and elimination of memorizing passwords or carrying tokens. The threat of breaching the security of the confidential data motivates the development of the data hiding techniques in this paper. This paper contributes in enhancing the security of biometric systems by incorporating the concept of wavelet decomposition along with the fusion of biometric traits. The concept of wavelet decomposition of feature templates helps in reduction of template size as well as it increases the compatibility of the templates of different biometric traits. The biometric key is generated from a biometric construct using proposed cryptographic key extraction algorithm and then the key is applied on fused template to protect the template from various attacks. The implementation results obtained provides 100% GAR at 17% FAR i.e. authentication performance of the system is better as compared to other systems.

Keywords: Authentication · Biometric encryption/decryption
Biometric template protection · Cryptography · Wavelet

1 Introduction

Data security stresses over the certification of secrecy, trustworthiness and accessibility of personal information [1, 2]. Biometric is one of the advancements utilizing the exceptional behavioral or physical components of a person to identify the distinguished user [3]. Utilizing biometrics to authenticate human is easy to use, demands less cost and offers better security measures to maintain a strategic distance from information theft and security provocation [4]. Unimodal biometric Systems are developed to get privacy and security but it is highly influenced by different attacks like function creep, intrusion attacks, etc. Hence, use of multiple biometrics (e.g., Fingerprint, Iris and face) together is generally utilized as a part of some large scale biometric applications (e.g., FBI-IAFIS). It is beneficial as compare to single biometric system as it provides lower error rate, enhanced accessibility, a higher level of flexibility, and less susceptible to spoof attacks. Cryptography is the art of utilizing mathematical concepts to encode or translate the original biometric information [5]. It permits the storage of confidential information on an unreliable system like the web so that nobody can fetch it without the permission.

Z. Patel and S. Gupta (Eds.): ICFITT 2017, LNICST 220, pp. 264–273, 2018.
https://doi.org/10.1007/978-3-319-73712-6_28

Multi-biometric Cryptosystem is the science and innovation of deciding and quantitatively assessing various biological information particularly utilized for authentication purposes. Physical as well as behavioural biometric features are acquired from accurate sensors and individual features are extracted to form a biometric template for the enrolment process. At the time of identification or authentication, the system processes another biometric input which is compared against the stored templates yielding acceptance or rejection of the user. This system is commonly utilized for reducing misuse and storage of the private biometric templates which offers an advanced solution for the era of the cryptographic key, encryption procedure and protection of the biometric templates [6].

In this framework, unique biometric templates are changed into biometric- subordinate data which helps in recovering cryptographic keys [7, 8]. Matching is specifically performed by confirming the legitimacy of reconstructed keys.

The method based on multiple fingerprint is first proposed by Soutar [9]. In enrolment stage, unique features from the acquired biometrics are extracted. Correlation function between each individual input is calculated by applying inverse Fourier transform on the product of applied biometric inputs. Technique is advantageous due to ease of implementation but it has very poor accuracy.

Fuzzy commitment scheme is a combination of error correcting code and cryptography to achieve cryptographic primitive proposed by Juels and Wattenberg [10]. This method is advantageous because of good accuracy but it is not able to perform well with multi-modal biometrics.

Table 1. Comparison analysis on different techniques

Techniques	Author	Year	Char.	FAR/FRR
Biometric encryption	Soutar [9]	1998	Iris	0.03/0.054
Fuzzy commitment	Juels & Wattenberg [10]	1999	Iris	0.47/0
Fuzzy vault	Juels & Sudan [11]	2002	Fingerprint	5/0.01
Quantization	Feng & Wah [12]	2003	Online signature	28/1.2
Bio-hashing	Teoh [13]	2006	Face	0.93/0
Shielding function	Tuels [14]	2003	Fingerprint	0.054/0.03
Hybrid template protection	Y.J. Chin & T.S. Ong [15]	2014	Palmprint & Fingerprint	1/0

Fuzzy vault scheme uses cryptography construction proposed by Juels and Sudan [11], designed to work with unique features from multiple biometrics e.g. iris pattern, minutiae from fingerprints, etc. In this method, features are represented as an unordered set. Main advantage of this method is that due to the variation in biometric data at authentication side, the ability of the biometrics to work with an unordered set of Fuzzy vault scheme provides the promising solution to improve the security [16–18] but disadvantage is that security of this technique decreases because of infeasible reconstruction of the polynomial generated from the Reed-Solomon code. For the authentication of online signature, Shielding function is proposed by Tuels [19, 20] but it does not work well on multiple-biometric template.

2 Proposed Framework for Multi-biometric Cryptosystem

The paper proposes a technique to implement a framework for Wavelet- decomposition based feature level fusion for Multi-biometric cryptosystem as shown in Fig. 1. The system comprises of two basic modules: (1) Multi-biometric fusion and (2) Private template protection. Our aim is to reduce the FAR and FRR of the system as low as possible with optimum EER. The proposed system is implemented by using the concept of wavelet decomposition for fusion process and normal encryption algorithm is used for generation of the key. Wavelet decomposition is applied to each extracted feature templates and approximate coefficients are taken from each individual template to fuse multiple biometric templates and the key is added to make the fused template more secure. Single secure sketch is stored in the system database. Whenever, user comes for the authentication, system requires enrolled biometrics and it will compare with the stored database. If system generates the same key which was used at enrollment process then the user will be genuine otherwise it will reject the user to access the system. System is divided in individual block and all blocks are described below

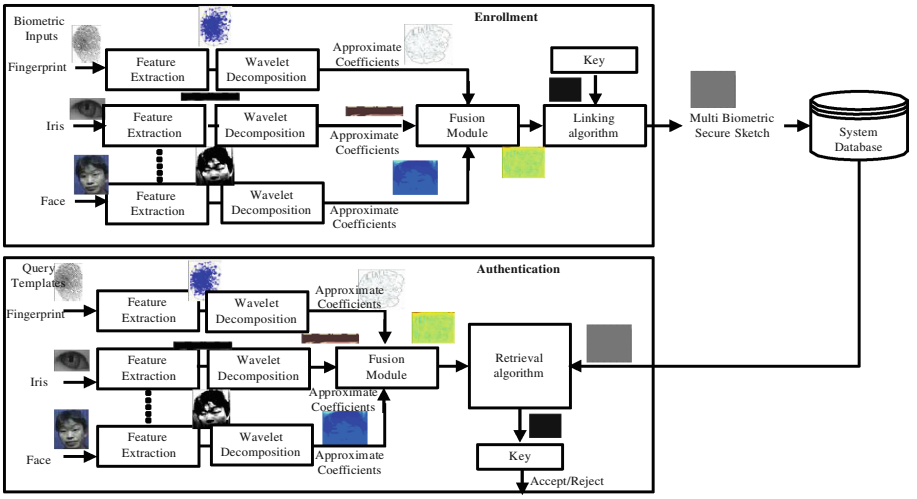


Fig. 1. Proposed multi-biometric cryptosystem using wavelet decomposition approach

2.1 Feature Extraction

Feature extraction methods for fingerprint, iris and face are described below.

Fingerprint feature extraction
It will extract ridges and bifurcation i.e. minutiae points from the fingerprint images using fingerprint image enhancement algorithm [21]. This technique requires less preprocessing and works better even with low-quality images.

Iris feature extraction
The image of iris includes undesirable data such as the pupil, sclera, eyelid, and eyelashes. Hence, before feature extraction, it must be pre-processed to remove unwanted portion. So, the preprocessing unit for the iris is required to perform image enhancement, iris segmentation, and iris normalization [22].

Face feature extraction
An important feature of the face extracted by applying 2D Principal Component Analysis (PCA) on the original biometric template. Suppose that there are training samples of images of $m \times n$ size, then the covariance matrix is calculated by

$$C = \frac{1}{N}\sum_{j=1}^{N}\left(A_j - \overline{A}\right)\left(A_j - \overline{A}\right)^T \tag{3}$$

Where, A is the sample image and $\overline{A}$ is a mean of sample images. The eigenvectors is calculated from the covariance matrix. The Eigen decomposition needs to be obtained by applying Singular Value Decomposition (SVD) then the data is simply projected onto the largest eigenvectors. To reduce the dimensionality, let V be the matrix whose columns contain the largest eigenvectors and D be the original biometric data. Then the projected data D′ is obtained as $D' = V^T D$ [23].

2.2 Wavelet Decomposition

Wavelet decomposition provides the decomposition matrix of applied feature template. The approximation and detailed coefficients are extracted from the decomposition matrix and approximate coefficient is taken for further processing of data because detailed decomposition matrix reduces the visibility of the biometric template [24]. In this paper, a single level of decomposition is used. Here, low-pass approximation coefficients and high-pass detailed coefficients are extracted. By using n level of decomposition the wavelet decomposition produced2n different sets of coefficients. Due to the down-sampling, the number of coefficients produced by decomposition process is same and there is no redundancy present.

2.3 Fusion of Feature Templates

Fusion of templates in the biometric system is not only the solution to the problem of single biometric, but it enhances the matching accuracy of the system. In order to extract relevant features and to remove the unwanted noise from raw biometrics, pre-processing and feature extraction is performed prior to fusion which overcomes the weakness of decision level fusion [25]. It is formed when feature vectors generated by multiple biometric templates are fused as a unified entity [25–27] so it has a prerequisite to first identify their nature and then apply only suitable algorithm to the biometrics. Here, the fusion is done by finding the contribution parameter (c_i) from each feature template $I(x, y)$ i.e. $c_i = \text{mean}\ (\text{mean}\ (I(x, y)))$ multiply by approximate

coefficient produced by wavelet decomposition method i.e. $I_A(x, y)$ [27]. Mathematically, it is represented by

$$f(x, y) = \sum_{i=1}^{n} c_i * I_A(x, y) \tag{4}$$

2.4 Key Generation Technique

Keys are directly generated from the fused feature templates. In existing methods, key is generated by applying random matrix and linear block coding on fused template i.e. RS encoder, BCH Encoder, etc. Hence, it is easy to detect by intruder. To improve the security of the system, instead of using one concept, multiple concepts are used to generate the key in proposed algorithm.

1. Generate the random matrix (R) of size of fused template.
2. Multiply the randomize matrix (R) by fused template (f).

$$\text{i.e.} \quad I_r = R(x, y) * f(x, y) \tag{5}$$

3. Perform the transform order (α) along the fused template [34] i.e.

$$\alpha = \frac{1}{M_1}\left[\left(\sum_{i=0}^{length(f)} f(i)\right) mod\left(2^L - 1\right)\right] \tag{6}$$

   Where, M_1 is the length of fused template and L is the number of bits used to represent the fused template.
4. Perform Hessenberg decomposition on the I_r to get orthogonal matrix Q_1.
5. The encrypted image is generated by

$$I_e = Q_1 * \alpha * Q_1' \tag{7}$$

6. Perform wavelet decomposition on I_e and find the approximate coefficient I_A. The key is generated by $k = I_A(x, y)$.

2.5 Key Binding Technique

Binding of the key (W_i) and fused template improve the privacy of feature templates in the system so that intruder cannot directly attack on the fused template. It is done by finding the average between fused template and generated key

$$W = avg\,(k(x, y), f(x, y)) \tag{8}$$

W is stored as a hash function in the system for further processing the data.

2.6 Key Retrieval Technique

By applying query templates at authentication side, the system calls the database template which is stored in secured form W_i to retrieve the key which is used to register the biometric templates. So it is the process between function of Hash value and query templates f' applied to verify the user. It is represented by

$$k' = 2 * W(x, y) - f'(x, y) \tag{9}$$

3 Experimental Results and Discussion

The fingerprint database for the design of the system is obtained from FVC 2004 while iris and face databases are obtained from CASIA. The experiment has been carried out in 4 set with 50 users in each technique. In set 1, the fusion of multi- fingerprint has been utilized. In set 2, the fusion of multi-iris has been utilized. Similarly, In set 3, a fusion of fingerprint and iris have been utilized. In set 4, a fusion of fingerprint, iris and face have been utilized. The ROC curves using different algorithm of fusion and template protection method is shown in below figures.

It shows the fusion of multi-biometric done using wavelet decomposition approach and Fourier transform of the concatenation of each template. Template protection methods use RS encoder and encryption algorithm to generate the key.

The parameter analysis of this cryptosystems is done by False Acceptance Rate (FAR) - It is the probability of an imposter being accepted as an authorized user, False Rejection Rate (FRR) - It is the probability of a legitimate user being rejected as an imposter, Equal Error Rate (EER) – It is the rate on which equal FAR and FRR is achieved, Genuine Acceptance Rate (GAR) – It is the rate at which the correct information is retrieved by the genuine user. Biometric cryptosystems require the helper data so even if it is attacked also it won't reveal the original information [7].

Table 1 shows the comparative performance analysis of different techniques proposed by different authors for multi-biometric cryptosystem.

Figure 2 shows that fusion is done by using wavelet decomposition approach and RS encoder is used to generate the key as a template protection method. It is observed that there is 50% EER with very high threshold i.e. very poor accuracy and GAR is also poor. So, we modified the fusion strategy by using Fourier transform of the concatenation of each template and encryption algorithm to generate the key as a template protection method. It is observed that there is 20% EER with low threshold i.e. very good accuracy and GAR is also 100% on FAR 57% as shown in Fig. 3.

Figures 4, 5 and 6 show that fusion is done by using wavelet decomposition approach and encryption algorithm is used on derived fusion template to generate the key as a template protection method for Multimodal, multi-finger, multi-iris respectively. It is observed that there is 0% EER with very low threshold i.e. very good accuracy and GAR is also increasing with 17% FAR i.e. maximum security for Multimodal case.

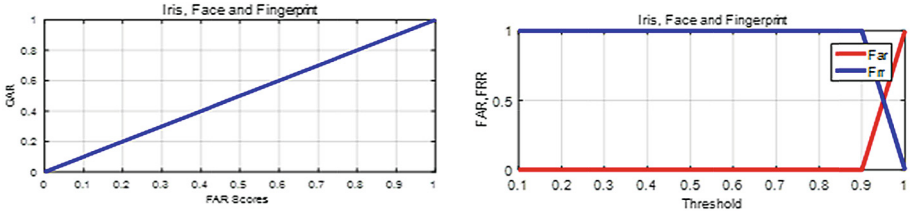


Fig. 2. Combination of wavelet decomposition on fusion and RS encoder as a key

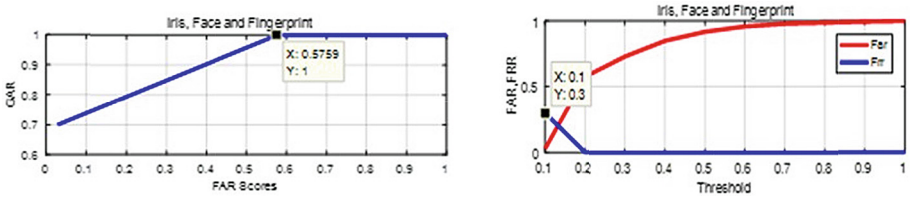


Fig. 3. Combination of fourier transform on concatenation of feature template and encryption algorithm on derived fused template

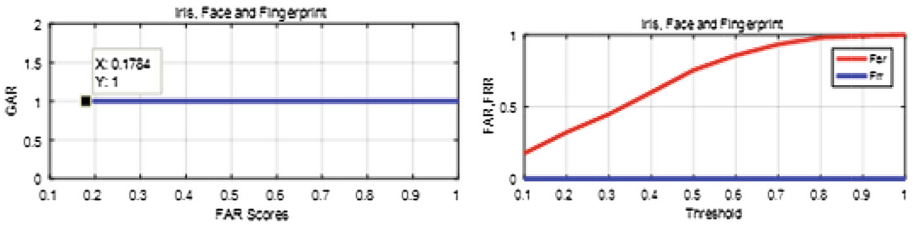


Fig. 4. Combination of wavelet decomposition on extracted feature in fusion process and encryption algorithm on derived fused template as a key based on fingerprint, Iris and Face

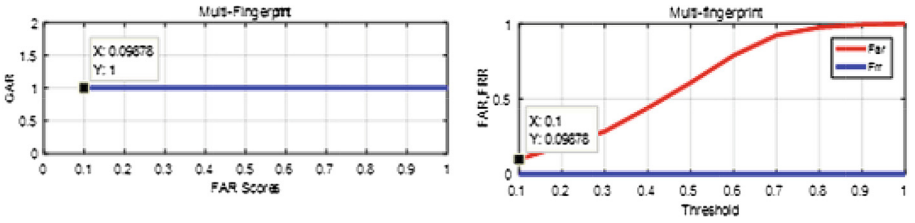


Fig. 5. Combination based on multi-fingerprint

Results show that proposed method gives better performance and security than existing technique. The hybrid approach gives 28% EER and 46% FAR with 100% GAR (threshold rate 13%) based on a combination of Fingerprint, Iris, and Face.

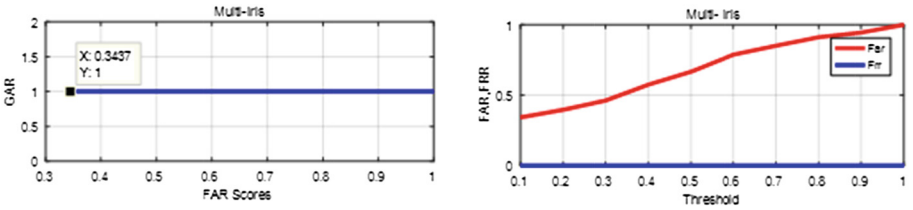


Fig. 6. Combination based on multi-iris

Wavelet decomposition approach gives 0% EER and 17% FAR with 100% GAR (threshold rate 0%). Here, threshold indicates the security levels. Lower threshold level indicates more security. Hence, wavelet decomposition approach provides lower FAR, FRR and better security on multi-modal biometrics. Proposed method along with existing techniques implemented using database mentioned above.

Tables 2, 3 and 4 show the comparison analysis on different techniques using multiple sets of biometrics.

Table 2. Comparison based on different fusion and private template protection technique

	Normal encryption [16]	Fuzzy commitment [17]	Fuzzy vault [18]	Shielding function [23]	Hybrid method [35]	Proposed method
Multi-finger	0.49/0.58	0.46/0.25	0.50/0.82	0.56/0.53	0.35/0.23	Error rate 0% with minimum threshold level i.e. maximum security
Multi-iris	0.50/0.70	0.42/0.59	0.43/0.48	0.42/0.26	0.5/0.63	
Finger & Iris	0.5/0.14	0.5/0.77	–	0.38/0.17	–	
Finger, Iris & Face		0.50/0.36	0.32/0.11	0.52/0.36	0.28/0.13	

Table 3. Comparison of different techniques with different combination of biometric trait in terms of EER/threshold

Fusion		Template protection technique (key)		Performance based on Iris, fingerprint, and face	
Concatenation [19]	Wavelet decomposition approach [24]	RS encoder [10]	Encryption algorithm [28]	EER/th	FAR/GAR
✓		✓		0.28/0.13	0.46/1
✓			✓	0.20/0.13	0.57/1
	✓	✓		0.5/0.95	1/1
	✓		✓	0	0.17/1

Table 4. Comparison of different techniques with different combination of biometric trait in terms of FAR/GAR

	Normal encryption [9]	Fuzzy commitment [10]	Fuzzy vault [11]	Shielding function [19]	Hybrid method [29]	Proposed method
Multi Finger	0.46/0.48	0.46/0.54	0.46/0.47	0.46/0.49	0.46/0.88	0.07/1
Multi Iris	0.46/0.47	0.46/0	0.46/0	0.46/0.66	046/0.46	0.34/1
Finger & Iris	0.46/0.46	0.46/055	0.46/0.93	0.46/0.75	0.46/1	0.7/1
Finger, Iris & Face	0.46/0	0.46/0.44	0.46/0.92	0.46/0.40	0.46/1	0.17/1

4 Conclusion

The proposed method developed multi-biometric system using the wavelet decomposition based fusion and Encryption algorithm on fused template as a template protection method that is accurate, reliable and secured. The proposed algorithm provides almost 0% EER and 100% GAR with 17% FAR on multimodal biometric and 0.9% FAR on same modal of biometric i.e. multi fingerprint. It is efficient compared to existing methods in terms of FAR, FRR and GAR.

References

1. Hellman, M., Deffie, W.: New direction in cryptography. IEEE Trans. Inf. Theory **22**(6), 644–654 (1976)
2. Devi, T.: Importance of cryptography in network security. In: International Conference on Communication Systems and Network Technologies (CSNT), April 2013
3. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(1), 4–20 (2004)
4. Uludag, U., Pankanti, S., Prabhakar, S., Jain, A.: Biometric cryptosystems: issues and challenges. Proc. IEEE **92**(6), 948–960 (2004)
5. Fu, B., Yang, S.X., Li, J., Hu, J.: Multibiometric cryptosystem: model structure and performance analysis. IEEE Trans. Inf. Forensics Secur. **4**(4), 867–882 (2009)
6. Wild, P., Radu, P., Chen, L., Ferryman, J.: Towards anomaly detection for increased security in multibiometric systems: spoofing-resistant 1-median fusion eliminating outliers. In: IEEE International Joint Conference on Biometrics (IJCB), September 2014
7. Jain, A., Nandakumar, K.: Biometric authentication: system security and user privacy. IEEE Comput. Soc. **45**(11), 87–92 (2012)
8. Toli, C.-A., Preneel, B.: A survey on multimodal biometrics and the protection of their templates. In: Camenisch, J., Fischer-Hübner, S., Hansen, M. (eds.) Privacy and Identity 2014. IFIP AICT, vol. 457, pp. 169–184. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18621-4_12

9. Soutar, C., Roberge, D., Stoianov, A., Gilroy, R., Vijaya Kumar, B.V.K.: Biometric encryption using image processing (1998)
10. Juels, A., Wattenberg, M.: A fuzzy commitment scheme. In: ACM Conference on Computer and Communications Security, December 1999
11. Juels, A., Sudan, M.: A fuzzy vault scheme. J. Designs Codes Cryptogr. **38**, 237–257 (2006). Springer
12. Feng, H., Wah, C.C.: Private key generation from on-line handwritten signatures. Inf. Manag. Comput. Secur. **10**, 159–164 (2002)
13. Teoh, A., Kim, J.: Secure biometric template protection in fuzzy commitment scheme. IEICE Electron. Express **4**, 724–730 (2007)
14. Huixian, L., Man, W., Liaojun, P., Weidong, Z.: Key binding based on biometric shielding functions, August 2009
15. http://www.scholarpedia.org/article/Cancelable_biometrics
16. Moon, D., Choi, W., Moon, K., Chung, Y.: Fuzzy fingerprint vault using multiple polynomials. In: IEEE 13th International Symposium on Consumer Electronics, May 2009
17. Jain, A.K., Pankanti, S., Nandakumar, K.: Fingerprint-based fuzzy vault: implementation and performance. IEEE Trans. Inf. Forensics Secur. **2**(4), 744–757 (2007)
18. Yang, X., Cao, K., Tao, X., Wang, R., Tian, J., Li, P.: An alignment-free fingerprint cryptosystem based on fuzzy vault scheme. J. Netw. Comput. Appl. **33**(3), 207–220 (2010)
19. Uhl, A., Rathgeb, C.: A survey on biometric cryptosystems and cancelable biometrics. J. Inf. Secur., January 2011. Springer International Publishing
20. Chikkerur, S., Cartwright, A.N., Govindaraju, V.: Fingerprint enhancement using STFT analysis. J. Pattern Recogn. **40**(1), 198–211 (2007). Elsevier
21. Hong, L., Wan, Y., Jain, A.: Fingerprint image enhancement: algorithm and performance evaluation. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 777–789 (1998)
22. Zaveri, M., Kapur, A., Gawande, U.: A novel algorithm for feature level fusion using SVM. In Hindawi Publishing Corporation (2013)
23. Tokumoto, T., Lee, M,. Ozawa, S., Choi, Y.: Incremental two dimensional two directional principal component analysis for face recognition. In: IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011) (2011)
24. Pavithra, C., Bhargavi, S.: Fusion of two images based on Wavelet transform. Int. J. Innov. Res. Sci. Eng. Technol. **2**(5), 1814–1819 (2013)
25. Kaur, H., Rani, E.J.: Analytical comparison of various image fusion techniques. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **5**(5), 442–448 (2015)
26. Purushotham, A., Usha Rani, G., Naik, S.: Image fusion using DWT & PCA. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **5**(4) (2015)
27. Ramli, D.A., Jaafar, H.: A review of multibiometric system with fusion strategies and weighting factor. Int. J. Comput. Sci. Eng. **2**(4), 158–165 (2014)
28. Bhatnagar, G., Wu, Q.M.J.: Biometric inspired multimedia encryption based on dual parameter fractional fourier transform. IEEE Trans. Syst. Man Cybern. **44**(9), 1234–1242 (2014)
29. Patel, H., Paunwala, C., Vora, A.: Hybrid feature level approach for multi- biometric cryptosystem. In: IEEE International Conference on Wireless Communications, Signal Processing and Networking (Wispnet-2016), Chennai, 23–25 March (2016)

Author Index