1st International Conference on Applied Soft Computing Techniques 22 & 23.04.2017 In association with International Journal of Scientific Research in Science and Technology

VOLUME 3, ISSUE 5, MARCH-APRIL-2017

ICASCT '17

1st International Conference on "APPLIED SOFT COMPUTING TECHNIQUES"

22-23 April 2017

Sponsored by

Council of Scientific and Industrial Research (CSIR)

Organized by

Department of Computer Applications

Kalasalingam University

Kalasalingam Academy of Research and Education

Ananad Nagar, Krishnankoil 626126 Virudhunagar District, Tamil Nadu India

In Association with

International Journal of Scientific Research in Science and Technology

International Journal of Scientific Research in Science and Technology

<u>About the University</u>

Kalasalingam University, formerly known as Arulmigu Kalasalingam College of Engineering (AKCE) was started in 1984. Late "Kalvivallal" Thiru T. Kalasalingam, a veteran freedom fighter is the founder Chairman of the College. The Ministry of Human Resources Development, Government of India, New Delhi has granted University status (Under section 3 of UGC Act 1956) wide notification No. F.9-2/2002/U.3 dated 20th October 2006 to AKCE. The University is located at the foot hills of the Western Ghats in a rural setting of Virudhunagar District. It started functioning from the academic year 2007-08. The University is located at the foot hills of the Western Ghats in a rural setting of Virudhunagar District. At present the university is offering 28 Under Graduate, 44 Post Graduate Programmes and doctoral programs. Kalasalingam University is the first institution in Asia to offer the six year Integrated B.Tech. programme for the Speech & Hearing Impaired Persons (SHIP). The university is accredited by NAAC with A grade and 3 of the U.G. programmes are accredited by NBA.

TIFAC-CORE in Network Engineering and National centre for Advanced Research in Discrete Mathematics (n-CARDMATH) has been set up with the financial support of DST, Government of India to promote research. The International Research Centre (IRC) at KLU is a testament to the vision of the founder to create a true hub of research that goes beyond the boundaries of the country. Over 1 lakh square feet is dedicated to the development of research and innovation in the fields of computing, energy and materials. The university has signed MOU with foreign universities, research organizations and industries with an objective of promoting research and development

About the Department

The programs are designed to cater to ever-changing needs and demands of the IT industry. The department aims to provide the best information and communication technology based solutions and value added services. The focus is on holistic learning to help the students to make significant contributions to all levels and meet the expectations of stakeholders. The department has committed members of faculty. The faculty members are highly motivated and devoted in delivering the highest quality professional education to students and strive to excel in their research areas. The placement record of the department had always been impressive.

Infrastructure with Industries

The department has close tie up with several software industries including Oracle, HCL, eSrit Technologies, HSL software. To Many consultancy projects were carried out to provide to the needs of industries such as Tamilnadu Cements, TVS, Satyam, Ramco Systems etc.

Infrastructure Facilities

Our campus is WIFI enabled campus. In our department more than 200PC's connected in LAN with advanced servers under windows and Unix environments is established. Switch based network for speedy internet access is available. All nodes are connected with 146 Mbps dedicated bandwidth. Round the clock lab access is provided. Special software for neural networks, image processing, fuzzy logic etc is also available.

Student Activities

Our students have developed software which helps in the learning process of the hearing impaired people. Our students have also developed software's for Exam Administration System (EASY), Attendance Information Systems (AIMS), Student Information System (SIS), Staff Attendance and Leave Transaction (SALT), Students SMS (S-SMS), Online Registration System (OCRS) Hostel Management System, Finance Management System, Health Center Management and Biometric Based Attendance Monitoring System for staff.

This software's are currently used in our University. Moreover our students have also developed school automation software as per the needs of schools. Competitions Students are motivated to participate in department association activities like quiz, paper presentation etc. Moreover they are allowed to participate in inter and intra collegiate to equip them with today's competitive world.

About the International Conference

The "**International Conference on Applied Soft Computing Techniques"** will enable the faculty member, research scholar and postgraduate students to gain experience in their research work related to applied computing techniques. Soft Computing being very important field in the advancement of technology is essential to be known by all faculty member research scholars and postgraduate students' atleast at a maximum level. This conference provides an opportunity to understand and explore various soft computing techniques in detail. The conference will also be an exciting and beneficial platform not only for networking among other faculty members, research scholars and postgraduate students but also for building the essential skills needed to venture into the career world. Eminent Speakers from reputed institutions and industries are invited for the seminar for delivering ideas and lectures on this emerging field.

From The Chancellor's Desk

"Illayavallal" Dr K Sridharan

Chancellor

Kalasalingam University

Our institution is providing a wide and varied arena for the staff and student communities to showcase their academic, research and extracurricular talents.

We believe in the absolute development of our students, research scholars, staff members and encourage them by providing the best of incentives and infrastructure. Students and research scholars are expected to broaden their horizons academically too, and the pursuit of knowledge over and above the prescribed. Our staffs are adequately equipped to guide students in the right direction in all fields of study and co-curricular activities too.

I am proud to say that once our students step in, they step out with self-confidence and Leadership knowledge to face all endeavors with complete savvy. The students and researchers are trained to have a sound foundation in the fundamentals of computer applications and soft computing techniques, a high level of practical skill in the use of that technology and at the same time, are sensitive to the issues prevailing in the society.

"ICASCT'17" will act as platform for exposing the talents of faculty members, research scholars and students.

Wish you all success!

Dr.K.Sridharan
Chancellor

From The Director's Desk

Dr S ShasiAnand

Director

Kalasalingam University

Kalasalingam University is recognized in the society for outstanding performance of its constituent units. Owing to the credentials of this college, the National Assessment and Academic Council (NAAC) and National Board of Accreditation (NBA) have assessed and accredited to grade 'A'. The college has been consistently attracting students from all parts of the country and abroad for its academic excellence .It is aimed at overall development of the students.

In order to accomplish our vision and mission, we are prepared the students and research scholars to take as much effort as possible for the betterment of academic scenario in India. We believe that education is an effective medium of social transformation. We get encouragement, looking at bright and successful careers of our students and research scholars, which subsequently benefit the society.

I am glad to know that the Department of Computer Applications is organizing International Conference on Applied Soft Computing Techniques –ICASCT '17 to cultivate the talents of researchers and PG scholars to face this competitive world. I appreciate the students and staff members for their untiring efforts in successful conduction of this International Conference.

Dr.S.ShasiAnand
Director

Dr.S.Saravanasankar,

Vice Chancellor

Kalasalingam University

Kalasalingam University provides the best learning environment in order to achieve international standards and prepare the students to a level where they can handle any situation independently to achieve higher degree of excellence in professionalism.. We intend to develop it as a lively and center of learning and research.

The University has a well-qualified, experienced and dedicated faculty. The highly pragmatic and progressive environment developed by the talented faculty is conducive for the development of learning, empowerment of knowledge and skill enhancement. University believes in providing close to life experience based teaching and learning techniques.

I am happy to note that the Department of Computer Applications is successfully organizing the International Conference on Applied Soft Computing Techniques –ICASCT '17 to cover the modern technologies in modern soft computing techniques. International conference is a way through which the hidden symbols of proficiency are posed to others
I trust "ICASCT'17 will act as a gateway for such an exposure.

Dr.S.Saravanasankar
Vice Chancellor

From the Organizing Chair Desk

Dr.D.Amutha Guka,

Head of the Department,

Department of Computer Applications,

Kalasalingam University, Krishnankoil

On behalf of the Organizing Committee, I am honored and delighted to welcome all the participants for the First International Conference on Applied Soft Computing Techniques (ICASCT'17) which is technically Co-Sponsored by CSIR, India. Under the guidance of our management and continues stride gave the way of success with confidence. The sharp, clear sighted vision and precise decision making power of our management has benefitted to stay competitive.

The main objective of this conference is to provide a concrete platform which will encourage and support scholars, researchers and industry professionals to carry and accomplish their research dreams in innovating new applications in applied soft computing techniques.

ICASCT-2017 was a great success where more than 70 Participants representing academic institutions presented their work. I hope all the participants will take advantage of this opportunity and contribute, through presentations, discussion and interaction to the development of new ideas and new directions in research and applied soft computing techniques.

As a final note, I would like to take this opportunity to thank all individuals who had contributed to this conference, including authors, keynotes speakers, the reviewers, the members of all committees, and sponsors.

Dr.D.Amutha Guka
Organizing Chair

From the Convenors Desk

Dr.S.RamKumar and S.Kartheeswaran
Assistant Professors,
Department of Computer Applications,
Kalasalingam University, Krishnankoil

It is a matter of great pleasure to see our department organizing its first International Conference on Applied Soft Computing Techniques**.** We heartily welcome all faculty members, research scholars, postgraduate students, industry members and the participants from various institution and industries to this conference.

It is very much heartening to see the immense response received by the conference from the research community. The main objective of this conference is to provide an opportunity to understand and explore various soft computing techniques in detail.

The conference will also be an exciting and beneficial platform not only for networking among other faculty members, research scholars and postgraduate students but also for building the essential skills needed to venture into the career world. We would like to acknowledge all the organizing team members. Without their outstanding job and planning, we would not have such an excellence conference.

We wish that ICASCT'17 will keep on growing in coming years with more impact on the international research community.
Expecting a sound response from you all.

CONVENOR(S)

ICASCT'17

Chief Patron (s)

"Sevarathna" Shri. Dr. K.Sridharan
Chancellor, Kalasalingam University.

Dr. S. ShasiAnand,
Director, Kalasalingam University.

Mr.S.ArjunKalasalingam,
Director , Kalasalingam University.

Patrons

Dr. S.Saravanasankar,
Vice-Chancellor, Kalasalingam University.

Dr.V.Vasudevan,
Registrar, Kalasalingam University.

Conference Chair

Dr.S.Arumugam,
Dean Research & Development, Kalasalingam University.

Organizing Chair

Dr.D.AmuthaGuka,
HOD,
Department of Computer Applications,
Kalasalingam University.

EDITOR-IN-CHIEF

Dr. Kun Ma,

Shandong Provincial Key Laboratory of Network Based Intelligent Computing

University of Jinan, China.

INTERNATIONAL ADVISORY BOARD

1. Prof. Xavier Fernando,

    Director, Ryerson Communications Lab. Ryerson University, Canada

2. Prof. Yiu-Wing Leung,

    Hong Kong Baptist University, Hong Kong

3. Prof. Sattar J. Aboud,

    Senior Adviser for Science and Technology of the Iraqi Council of representatives, Iraq

19. Dr.K.Satheesh Kumar, Dept of CS&IT, Kalasalingam University.

20. Dr. K.Sumathi, Dept of CS&IT, Kalasalingam University.

LIST OF KEYNOTE SPEAKERS

1. Dr. Nazri Bin Mohd Nawi, Dean, Faculty of CS& IT, Universiti Tun Hussein Onn Malaysia, Serdang.

2. Dr. K.V. Ramesh, Principal Scientist, CSIR- Fourth Paradigm Institute (CSIR 4PI), NAL Belur Campus, Bangalore - 560 037, India.

3. Dr. V. Rakesh, Scientist, CSIR- Fourth Paradigm Institute (CSIR 4PI), NAL Belur Campus, Bangalore - 560 037, India.

4. Dr.U.Srinivasulu Reddy, Dean, Student Welfare, Department of Computer Applications

   National Institute of Technology, Tiruchirappalli – 620 015.

5. Dr. Jayavel Sridhar, Assistant Professor, Department of Biotechnology, Madurai Kamaraj University, Madurai-625021, Tamilnadu, India

ORGANIZING MEMBERS

1. Dr. S.Balamurali, Dept of Computer Applications, Kalasalingam University.

2. Dr.K.Maheswari, Dept of Computer Applications, Kalasalingam University.

3. Mrs.G.A.Pethunachiyar, Dept of Computer Applications, Kalasalingam University.

4. Mr.M.Arun, Dept of Computer Applications, Kalasalingam University.

5. Mr.J.Pradeep Kandhasamy, Dept of Computer Applications, Kalasalingam University.

6. Ms P.Buvaneswari, Dept of Computer Applications, Kalasalingam University.

7. Ms. P.Packya Amutha Priya, Dept of Computer Applications, Kalasalingam University.

ORGANIZING SECRETARY

Dr. S.Ramkumar & S.Kartheeswaran

Department of Computer Applications,

Kalasalingam University

Ananad Nagar, Krishnankoil 626126

Virudhunagar Dist, Tamilnadu

Advisory/Editorial Board

Dr. Anant Lalchand Chaudhari, Department of Electronics, Arts, Science & Commerce College, Chopda, Jalgaon, Maharashtra India
Dr. N. Pughazendi, Computer Science and Engineering, Panimalar Engineering College Chennai, Tamilnadu, India
Dr. V. Ananthaswamy, Department of Mathematics, The Madura College, Madurai, Tamil Nadu, India
Rakesh K. Bumataria, Mechanical Engineering, Marwadi Education Foundation's Group of Institutions Rajkot, Gujarat, India
Dr. Arvind Bijalwan, Indian Institute of Forest Management (IIFM) (Ministry of Environment & Forests, Govt. of India) Bhopal, Madhya Pradesh, India
Sharvil D. Shah, Mechanical Engineering Dept. Parul Institute Of Engg. & Tech, Vadodara, Gujarat, India
Dr. Aditya Kishore Dash, Department of Environmental Engineering, Institute of Technical Education and Research (ITER), S'O'A University, Bhubaneswar, Odisha, India
Dr. Subha Ganguly, Department of Veterinary Microbiology Arawali Veterinary College, Bajor, Rajasthan, India
Dr. Shivakumar Singh, MVS Govt UG & PG College, Palamuru University, Mahabubnagr, Telangana, India
Md Irfan Ahmed, Power System, Sityog Institute Of Technology Aurangabad, Bihar , India
A. Dinesh Kumar, Mathematics, Dhanalakshmi Srinivasan Engineering College, Perambalur, Tamilnadu, India
Shyam Lal Sharma, Mechanical Engineering, Department, AFSET, Al Falah University, Dhauj, Faridabad, India
Prof (Dr.) Hardeep Singh, Electronics & Communication Engineering Department, Indo Global College of Engineering, Abhipur, District Mohali, Punjab, India
S. R. Boselin Prabhu, Anna University Chennai, Tamilnadu, India
N.R.Shingala, Department of Mechanical Engineering, VVP Engineering College, Rajkot, Gujarat, India
R. G. Vaghela, Mechanical Engineering, Atmiya Institute of Technology & Science, Rajkot, Gujarat, India
S. Mayakannan, Vidyaa Vikas College of Engineering & Technology, Tiruchengode, Tamil Nadu, India
R. Girimurugan, Nandha College of Technology, Erode, Tamil Nadu, India
Achal Garg, Structural Section, Keppel Offshore and Marine Engineering, Mumbai, India
Velladurai Chinnappillai, Department of Animal Repoduction, Gynaecology and Obstetrics, Veterinary College And Research Institute, Tamil Nadu Veterinary and Animal Sciences University, Namakkal, Tamil Nadu, India
Sachin Narendra Pardeshi, Department of Computer Engineering, R.C.Patel Institute of Technology, Shirpur, Maharashtra, India
Hari Krishna, Department of Chemistry, Jagans College of Engineering & Technology, Choutapalem, Vikrama Simhapuri University, Nellore, Andhra Pradesh, India
Rajnish Kumar, Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Lucknow Campus, Uttar Pradesh, India
Dr. Meenakshi Sharma, Medical Biotechnology, Dr B.R. Ambedkar Centre for Biomedical Research, University of Delhi, Delhi, India
Dr. P. Vijayarengan, Department of Botany, Annamalai University, Annamalainagar, Tamil Nadu, India
Dr. Vishnu Narayan Mishra, Applied Mathematics and Humanities Department, Sardar Vallabhbhai National Institute of Technology, Ichchhanath Mahadev Dumas Road, Surat, Gujarat, India
Dr. Bangole Narendra Kumar, Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India

International Advisory/Editorial Board

Md. Kamrul Hasan, English Language Institute, United International University Universiti Utara Malaysia, Malaysia
Dr. Eng. Ramzi R .Barwari, Department of Mechanical Engineering, College of Engineering, Salahaddin University - Hawler (SUH), Erbil - Kurdistan, Iraq
Kerorsa, Lubo Teferi [Environmental Law and Governance], Seoul National University; Family Dormitory. Seoul, South Korea
Dr. C. Viswanatha, Department of Chemistry, Arba Minch University, Arba Minch, Ethiopia
Tsunatu Danlami Yavini, Chemistry Department, Faculty Of Science, Taraba State University, Jalingo , Taraba State, Nigeria
Bello Alhaji Buhari , Usmanu Danfodiyo University, Department of Mathematics, Computer Science Unit, Sokoto, Nigeria
Ramzi Raphael Ibraheem AL Barwari, ANKAWA - ERBIL, Department of Mechanical Engineering, College of Engineering, Salahaddin University - Hawler (SUH), Erbil - Kurdistan
Innocent E. Bello, National Space Research And Development Agency, Abuja, Nigeria
Ang Kean Hua,Department of Environmental Sciences, Faculty of Environment Studies, Universiti Putra Malaysia, Selangor Darul Ehsan, Malaysia
Dr. Abdulghani Mohamed Ali Alsamarai, Professor of Infectious Diseases and Dermatology and Medical Education, Iraq
Agbor Reagan Bessong, Department of Genetics & Biotechnology, P.M.B 1115, University of Calabar, Calabar, Nigeria
Dr. Abbas Bubakar El-ta'alu , Department of Human Physiology, Faculty of Basic Medical Sciences, College of Health Sciences, Bayero University, Kano, P.M.B 3011, Kano-Nigeria
Dr. Salah H.R. Ali, Engineering & Surface Metrology Department, National Institute for Standards (NIS), Giza, Egypt
Jiban Shrestha, Plant Breeding & Genetics, Nepal Agricultural Research Council, National Maize Research Program, Rampur, Chitwan, Nepal
Dr. M. L. A. M Najimudeen, Department of Obstetrics and Gynaecology, Melaka Manipal Medical College, Malaysia
Rolando P. Javellonar, Romblon State University, Odiongan, Romblon, Philippines
Chaudhary Dhiraj Kumar, Department of Life Science, Kyonggi University, Suwon, Gyeonggi, South Korea

CONTENTS

A Survey on Medical Image Encryption

P. K. Kavitha*[1], P. Vidhya Saraswathi[2]
*[1]Research Scholar, Department of Computer Applications, Kalasalingam University, Krishnankoil, Tamil Nadu, India
pkkavitha78@gmail.com[1]
[2]Assistant Professor, Department of Computer Applications, Kalasalingam University, Krishnankoil, Tamil Nadu, India
vidhyasaraswathi.p@gmail.com[2]

ABSTRACT

In healthcare industry Medical images or information is transmitted from source to destination using wired or wireless medium. The transmission of this information requires more security. To transfer the Image information from one place to another encryption and decryption process is used. Encryption means converting data or information from its original form into converting form that hides the information in it. The Image data need to be protected from unauthorized access. Encryption is used to increase the data security. The encrypted Image is secured from any kind cryptanalysis. This paper provides Literature survey of various Image Encryption and Decryption Techniques, methods and algorithms for Medical images.

Keywords: Encryption, Decryption, Cryptanalysis

I. INTRODUCTION

Nowadays Internet is used for faster transmission of valuable information. These information may be text, image, video anything . The Internet has many types of attacks. So the valuable information needs to be protected from the eavesdroppers. Medical Images are used in various fields such as Medical Science, Military communications, Biometric field, Medical Imaging, Telemedicine, Online photograph album, etc. The security of the medical images is based on (i) Confidentiality (ii) Integrity (iii) Availability.
Confidentiality: The protection of data from unauthorized user. The sender and receiver understand the contents of the message which is transmitted.
Integrity: The content of their communication is not changed in transmission.
Availability: Timely accessibility of data to authorized entities.
To exchange the image securely over the network, the image encryption is used. Unauthorized user can't able to decrypt the image. Text Encryption varied from Image Encryption. The characteristics of image are big functionality, high redundancy and correlation between the pixels. Images are vast in size, already existing encryption methods are difficult to apply and sluggish in manner. To transfer digital data into a cipher code encryption process of mathematical algorithms and keys used. To get back the original image or text from converted form that is cipher data, decryption with the mathematical algorithms and keys used.

In image encryption method, we first study the differences between implementations for image data and text data. There are a number of differences between the image data and text data as follows. Text data are sequences of words. Text data can be encrypted directly by using block or stream ciphers. When the cipher text is produced, it must be decrypted to the plaintext in a lossless way. The cipher image can be decrypted to the plain image in some lossy manner. Digital images are represented as 2D arrays. To protect stored 2D data, they must be converted into a 1D array before using encryption techniques. Encryption and Decryption are used by image compression for reducing its storage space and transmission time.

II. CRYPTOGRAPHY TERMINOLOGY

Plaintext: The source message.

Cipher text: The distorted message.

Key: Critical statistics utilized by the Cipher, Only the Sender and Receiver recognizes the key.

Cipher: An algorithmic rule for remodelling Plaintext to Cipher text.

Code: An algorithmic rule for translating an understandable message into an meaningless message.

Encipher: (Encrypt) Converting Plaintext to Ciphertext the usage of a Cipher and a Key.

Decipher: (Decrypt) Converting Ciphertext into Plaintext the usage of a Cipher and a Key.

Cryptology: The combination of Cryptography and Cryptanalysis.

Cryptography: Study of Encryption Principles and strategies.

Cryptanalysis: (Code Breaking) The look at of standards and techniques of deciphering Ciphertext without knowing a key.

Hash algorithm: An algorithmic rule that converts textual content into a string of fixed length.

Secret Key Cryptography (SKC): Single key used for both enciphering and deciphering.

Public Key Cryptography (PKC): Two Keys used. One key for Encipher and some other for Decipher.

Public Key Infrastructure (PKI): PKI is a Certificate authority.

III. CRYPTOGRAPHIC ALGORITHMS

Depending upon the number of keys used, cryptographic algorithms can be categorized into two types :

- Symmetric algorithms (Secret Key)
- Asymmetric algorithms (Public Key)

Symmetric : It is also known as Secret key or Private key. Sender and Receiver used a single key for enciphering and deciphering. Examples : Data Encryption Standard (DES), Triple DES, and Advanced Encryption Standard (AES).

Asymmetric : Two different keys (public and private keys) are used for enciphering and deciphering. For Encipher, Public key is used. For Decipher, Private key is used. Examples : Rivest-Shamir-Adelman (RSA) and Elliptic Curve Cryptosystem (ECC).

Encryption Algorithm : Encipher message created as output. Enciphered data depends on the plaintext and the secret key. It performs various transformations and substitutions on the plaintext.

Decryption Algorithm : Reverse manner of Encipher. It takes the Cipher and the Secret Key as input and gives the plaintext as output.

An encryption algorithm for multimedia data based on arithmetic modulo 2. Both plaintext and ciphertext are of block size 64. Secret key is also of 64 bit length. The structure of this block cipher provide confusion and diffusion.

IV. LITERATURE SURVEY

A. **Q. N. Natsheh et. al [1]** have presented an algorithm using XOR Cipher with AES (Advanced Encryption Standard). DICOM is Digital Images and Communication in Medicine. This file contains Two Parts : Header Data and Client Data. Header Data(Textual Data) stores the Patient's Information. Clinical Data consists of Name, Scan Image Type, Pixel array attributes such as Pixel Depth etc. Pixel Data can be an image, short video or audio. DICOM offer confidentiality for header records. DICOM supports the huge kind of Digital Medical Images consisting of Computed Tomography, Magnetic Resonance Images (MRI). Encrypting Decrypting Medical images required greater computational time. In this paper, AES used for only one image is encrypted and XOR cipher for encrypting the remaining multi-frame DICOM images.

XOR Cipher : It is a symmetric encryption algorithm. XOR cipher is derived from Boolean Algebra XOR function that returns 'True' while two arguments have different values. XOR function can be applied to binary bits. In Enciphering, the strength of the XOR cipher depends on the Length and the nature of the Key. Lengthy Key achieves higher overall performance.

XOR-AES based Encryption:
First approach : First image in the Multi-frame DICOM images as XOR key to encrypt the rest of the images in the Multi-frames.
Second approach: To encrypt the key using AES with Counter(CTR) mode of operation.

This Algorithm is evaluated using computational time, normalized correlation, entropy, PSNR(Peak Signal to Noise Ratio) and Histogram analysis. Medical image confidentiality was achieved by using the XOR cipher. The XOR keys were generated randomly. The Encryption approach based on a random key that

provides better performance and shorter encryption, decryption time than Naive approach.

B. **Vratesh Kumar Kushwaha et. al [2]** have proposed a new technique that combines Encryption and Watermarking for protected transaction of medical image. In this scheme, ROI in the image as the watermark. Selected ROI portion is enciphered by linear feedback shift register totally stream ciphering. Again, this portion is encrypted with the public key. That key is derived from a Diffie-Hellman Algorithm. Region of Interest is embedded into the Medical Image by Spread Spectrum scheme.

Bit Plane Slicing : For analyze the image, keeping apart a digital image into bit planes. Every pixel is represented by using 8 bits. Suppose the image has eight 1-bit planes. It has the ranges from bit plane1 - 0(LSB) to bitplane7 (MSB). Bit Plane 0 includes all lowest order bits and Bit Plane 7 includes all high order bits. Bit-plane extraction for a 8-bit image, value of gray level transformation maps all levels between 0 and 127 to one level and maps all levels from 129 to 253 to another.

Diffie-Hellman : This is a way of exchanging cryptographic keys. It permits two persons that haven't any knowledge about each other. They shared their secret key through the Internet. That key is used to encrypt the Medical Image.

For Example, S and R are Sender and Receiver, then S and R agree a secret key. m and n are two large numbers m & n such that $1 < n < m$. S chooses random A and then computes $A = n^a \bmod m$. R chooses random B and then computes $B = n^b \bmod m$. S Computes Key1 $= n^a \bmod m$. R Computes Key2 $= n^b \bmod m$. Key1 = Key2 $= n^{ab} \bmod m$.

Watermarking : The process of embedding digital statistics into another for copyright protection, authentication and authorized verification. For Encipher, Use the equation $E = \sum(h, m, k)$ where E is the Stego Image, h is the Host Image, m is the watermark image and k is a secret key. For Deciphering, use the equation D(E, k).

Encryption
- Using Bit plane slicing, pick the MSB plane from the Medical Image
- Then, select the ROI from the MSB plane.

- Generate a 64-bit Secret key the use of LFSR and added to each pixel of the image.
- Using Diffie-Hellman algorithm, the Public Key is generated.

Decryption
- Encrypted watermark image chosen for extraction.
- Watermarked image deciphered by Stream cipher.
- Using Diffie-Hellman algorithm, Public key is generated.
- Add it to extract pixel of the encrypted image.
- Finally, the ROI image is acquired.

The Scheme turned into tested for Different Medical Images. That image may be MRI (Magnetic Resonance Imaging), CT-Scan (Computed Tomography) and X-Ray Images. In this paper, MRI Image of size 256 x 256 was taken as Cover Image and ROI of the image is taken as Watermark. The Public key is generated via using a D-H algorithm by growing the Level of Security.

C. **Simranjeet Kau r et. al [3]** introduced a new Reversible Data Hiding Technique for Authentication and Data Hiding. In this paper, ROI(Region of Interest) and NROI (Non Region of Interest) is defined. ROI is protected and efforts embed data in NROI. Here, Semi-reversible Scheme is capable of hiding patient's data. The Fragile Watermarking Technique is used to verify authenticity of the image, to achieve image authentication.

Watermark: Using Hash Function, generate a fixed hexadecimal number message to a particular message defined by the sender. Read the Text file. It contains Patient's information and Converted character into integer values.

Embedding the Watermark in NROI Steps
1. Read Image into MATLAB environment.
2. Convert it into Gray Scale Image.
3. Separate ROI and NROI using the Cropping Tool.
4. Read Diagnosis Report.
5. Generate watermark by combining step 3 and 4.
6. Put the integer form of concatenated character string data into an array called TABLE.
7. Scan the host image from TABLE and match for minimum difference match in NROI.
8. Confirm its location in a Secret Key array.

9. Update the Encrypted image array according to this newly found pixel and update the secret key.
10. The Watermarked signal image will be produced.

Extraction Process
- Load the Watermarked Image
- Extract the pixels by using the secret key in the sequence provided by secret key and put it in an array.
- Decrypt the extracted watermark.
- Compute the MAC code
- Compare the extracted hashes to the computed hash.
- If both are same, then received image is authentic.
- If both are not same, then received image is unauthentic.

To evaluate the performance DICOM image of brain of patient was used. Fragile Data Hiding Technique preserves the record of Medical Image through embedding the medical diagnosis report and other records. This approach lets in the storing and transmission of Electronic Patient Record beside with image authentication codes. The original image can be recovered perfectly. The scheme is good at authentication.

D. **C. Deepak Naidu et. al [4]** have presented a new algorithm to combine LSB(Least Significant Bit) algorithm with Blowfish algorithm. This is also known as combination of Steganography and Cryptography. Blowfish Algorithm is used to encipher and decipher of the image. LSB Algorithm is used to embed the message into the image. This method consists of three phases. There are Image Steganography, Cryptography and Decryption.

Image Steganography - LSB Algorithm.
The inputted image and the cover image are both covered into their binary equivalents. Each bit of the message is embedded into each pixel's LSB. This method keeps till all the bits of the message are embedded into the cover image. Then acquiring the Stego-Image.

Cryptography - Blowfish Algorithm.
Blowfish is a symmetric block cipher. It has a P-array and S-boxes. The P-array has 18 32-bit boxes such as P1,P2,P3,....,P18. S-boxes has 4 32-bit arrays such as S1,S2,S3,S4 with 256 entries each. This algorithm have 16 rounds.

Decryption
Encrypted image first decrypted using the Blowfish algorithm and then the hidden message can be obtained from the Stego-Image.
This paper mainly concentrates on the confidentiality of patient information during the transfer of Medical Images over the Internet. This algorithm is more efficient than many other algorithms. This method offers High quality images and the MSE (Mean Square Error) value is very less.

E. **A. Umamageshwari et.al [5]** affords a Novel Algorithms AHF (Additive Hash Function) and RSA. DS (Digital Signature) is used right here to attain high Confidentiality and Authentication. First, using JPEG2000 Medical Image is compressed. Then shared through open network.

JPEG2000 Image Compression : Four Steps in this process are the following
1. Pre-processing
2. Transformation
3. Quantization
4. Entropy Encoding
- **Pre-processing :** Input Medical Image is decomposed into components of most of 256. These additives are decomposed into rectangular tiles.
- **Transformation :** DWT(Discrete Wavelet Transform) used in JPEG2000. After pre-processing, each and every tile is decomposed into special resolution tiers. Resolution tiers are made from sub bands of coefficients.
- **Quantization :** Sub bands of coefficients are quantized. Also sub bands of coefficients are accrued as blocks.
- **Entropy Encoding :** Bit planes of the coefficients in code block are entropy encoded. ROI can be coded at a better pleasant than the historical past.

The Input image is an MRI image of size 512 x 512.

Additive Hash Function (AHF) : This accepts first row of the source image as input. To produce a constant length of output as a hash value (Message Digest), some confusion and diffusion methods applied mathematically. Output of hash value size may be 128 bits.

Procedure of AHF : Convert the image into 512 x 512 pixel. Then take the first row as another table. Divide 512 elements into 4 Divisions namely A1,A2,A3,A4. Each includes 128 elements.

Add Exchange Sets :

B1 = A1 + A3

B2 = A2 + A4

Subtract B1 and B2, M1024 = B2 - B1. Then Divide the M1024 into 8 elements. 16 Elements = 128 bits. c1,c2,c3,c4,c5,c6,c7,c8 are 8 elements.

Add Exchange values

M1 = c1 + c5   M2 = c2 + c6

M3 = c3 + c7   M4 = c4 + c8

Each value of M has 16 elements = 128 bits.

Add and Subtract interchange values of M.

HF1 = M3 − M1

HF2 = M4 + M2

Add HF1 and HF2 to get the hash value.

AHF = HF1 + HF2

Digital Signature using RSA Approach : Digital Signature is used for Authentication. DS is also used to verify the reliability of the source message. RSA method is used to generate the DS. Additive Hash Function accepts the Medical Image and gives 128 bit output of the hash value. The RSA algorithm is used to encrypt the hash value. Reversible watermarking used to extract the signature. Digital Signature is as compared with Extracted Signature. If they are equal, then original image isn't altered in some point of transmission.

Kerberos Algorithm : It is a secret key encryption scheme used for authentication. Tickets of authentication introduced to Kerberos. Medical professionals make use of a pair of keys for encipher. They used the system that contains no record of the username and password. The system launches a demand for the Kerberos Initial Ticketing Service for requesting a ticket yielding for the user. This request is unauthenticated. The Initial Ticketing Service creates an exclusive session key named Kusession and launch to the user.

Session key form :

|| Ttks.kusession| Ktks.Kusession|Kuser

The user decipher the TGT via his/her password as a key. If the decipher succeeds, the user is real.

Kerberos Ticket Form:

{TKT, {request, User ID, Time} Kusession}

where TGT = { Ttks,kusession} Ktks

Kerberos ticket yielding service uses its own secret key (Ktks) to decipher the demand it has expected. The session key (Kusession) deciphers the rest of the demand.

A strict authentication was achieved through Kerberos. This method successfully applied. Additive Hash Function and RSA method solves the dilemma of integrity, reliability and confirmation of Medical Image. A secret key is used for embedding and removal method. It provides further confirmation for Medical Images. It has high security.

F. **P. Antony Raj et.al [6]** have presented the JPEG Lossless algorithm for the purpose of compression and MD5 to compute the image to improve the authenticity hash value. Encryption process using Advanced Classical Cipher to shape the Digital Signature. First, DS and message are watermark. Then it is embedded in Digital image communication images.

Image Compression : Compression needed to avoid a collision between original image, watermark and Digital Signature. In this process, squashed representation of an image sinking the image storage and transmission requirements. Here, the compress Medical image before embedding JPEG Lossless using the DWT, so no need to block the image.

Digital Signature : Digital Signature computed using MD5 and ACC. Input Medical Image uses DS to verify the authentication. MD5 algorithm used to generate DS.

MD5 Algorithm : First accepts the input of any length and gives 128 bit stable output as the hash value. The hash value of the Medical Image is encrypted using ACC Algorithm. Patient details, Disease details and Digital Signature grouped as watermark. Sender side, using Reversible watermarking, watermark is fixed inside the image. Receiver side, Digital Signature and Patient, Disease details extracted from the image. The hash value of the input image is calculated.

Kerberos : Firewalls make a risky assumption that the attacker is coming from the outside for attacks frequently. Kerberos assumes that network connections are the weak link in network security. Using ACC, find the DS. Then the DS and extracted signature is

compared. If two signatures are equal, then no fluctuate in the image through transmission.

Authentication can be achieved from Web servers and Kerberos Technique. Secret key used for both Embedding and Extraction methods. It produces substantiation to Medical Images.

G. **Boukhatem Mohammed Belkaid [7]** presents a new Encryption method called as Hybrid Encryption for Secure transmission of Medical images. Encipher is done using AES - Advanced Encryption Standard and RSA - Rivest, Shamir and Leonard Adleman. AES is used for Data privacy. RSA is used for verification. AES is based on Substitution-Permutation network. RSA is a Public Key Cryptosystem. This is broadly used to protecting data transmitted. Here, Encryption key is public and Decryption key is set aside secret. Encryption generates a exclusive password to each newfangled session of encryption.

AES Algorithm : AES is also called as Rijndael. Substitution – Permutation network grouping is AES. It has a permanent block size of 128 bits and a key size of 128, 192 or 256 bits. Four steps in this algorithm are Substitute bytes, Shift Rows, MixColumns and AddRoundKey.

Block Cipher Operation : In Cryptography, mode of operation is an algorithm that uses a block cipher. A block cipher is appropriate for the safe communication of groups of bits called a block. Modes of operation is used for enhancing the Cryptographic algorithms. Five modes of operations are used are ECB(Electronic Code Book), CFB(Cipher Feedback), CBC(Cipher Block Chaining)and Counter(CTR). These modes are utilize with symmetric block cipher that is Triple DES and AES.

RSA Algorithm : RSA(Rivest Shamir Adleman) developed in 1977. RSA created by Ron Rivest, Adi Shamir and Leonard Adleman at MIT. The calligraphy RSA are the first letter of their surnames. It is a Public Key Cryptosystem. It is used for sheltered data broadcast. Here, Encryption Key is public. Decryption Key is set aside Secret. Keys based on two big prime numbers. The sender encrypts images using RSA algorithm. A public and private key ($Pub_E(b_x,n_x)$, $Priv_R(u_x,n_x)$) are used in emission. In reception, the public and private key ($Pub_R(b_y,n_y)$, $Priv_R(u_y,n_y)$) are used. The Sender encipher image using the key K with

the RSA. The private key of the $Priv_E$ to get a signed key K' such that : $K' = K^{ux} \bmod(n_x)$. K' is encipher next time using the RSA public key. Receiver make the key K". $K'' = K'^{by} \bmod(n_y)$.

AES technique reveals good characteristics by declining the correlation of adjacent pixels. Integrity is guaranteed by the correlation between adjacent pixels in the image. Several parameters used for various tests for analysis.

H. **Amarit Ma***mbutdee et.al [8]* proposes a new method of application to Medical Image Encryption and Watermarking to store Patient information using Scrambling Algorithm. Patient information is hidden. First, the most important detail of Patient will convert into Two-Dimension Barcode ECC200 standard. Then, Patients need to create six digit Password for Secure this image. The Patient's Password will be reused in Watermark Extraction step. For Watermarking, DCT - Discrete Cosine Transform and DWT - Discrete Wavelet Transform applied to block size of 8x8 pixels.

Discrete Wavelet Transform : It is a mathematical algorithm used for explaining the structure of signal by converting signals from time domain to frequency domain. DWT used in many areas like Digital Signal Processing, Image Processing, Image Compression, and also in researching of Digital Watermarking. It produces decomposes the given image into four groups of frequency sub-bands. That four groups are Low Resolution Approximation Component, Horizontal, Vertical and Diagonal. Embedding watermarking in low frequency sub-bands are more robust.

Discrete Cosine Transform : It is a very popular transform function. It is used to alter the signal from the spatial domain to frequency domain. DCT has used in Data Compression and Image Processing. Many researches transform images by DCT then divided into Non-overlapped m x m block.

Arnold Transform : It is a simple chaotic map. It is widely used for Image Scrambling by shifting the position of a pixel instead of change the value. To improve reliability of Scramble, Arnold Transform given a new parameters known as a, b. Here, a, b are real number call parameters. These both numbers are higher than 1 will increase stronger of chaotic.

Two Dimensional Bar Code(Data Matrix Barcode) : It was invented in mid of 1980s. It is used to eliminate the limit of the One-Dimension Barcode for data capacity and size. This is used for error checking and error correction algorithm. This is matrix barcode constructed as a square or rectangular symbol. The size or image depends on the amount of information. Data Matrix Barcode is composed of two separate parts : Finder Pattern and L Finding Pattern. Finding Pattern is used by the scanner to locate the symbol and the encoded data. The solid dark is called the " L Findig Pattern". It is used to determine the size, orientation and distortion of the symbol.

DWT, DCT coefficient values compared in each block for position identification of the embed message in Medical Image. This paper applies Scrambling process to increase high embed capacity with maintaining high PSNR(Peak Signal Noise Ratio value. For Security improvement, unique security key is used for patients.

I. Li-bo Zhang et.al [9] have offered a Medical image Encryption and Compression method. This is used for secure transmission of Medical image . New method encipher and compress the Medical Image by Compressive Sensing and Pixel Swapping based Permutation Technique. In Compressive Sensing, Source image is compressed and encrypted by Bernoulli measurement matrix.

CS with Cryptographic Features Embedded : It is worn for Sampling and Compression of signals. Take a Length L and signal S, it is said to be K-sparse if S can be approximated. In Compressive Sensing, the signal is measured through standard samples. Receiver side, Convex Optimization algorithms. Orthogonal Matching Pursuit (OMP) used for reconstruction. Compressive Sensing measurement Matrix uses Optimal Sensing Performance. This method gives Low Complexity and user friendliness.

Pixel Swapping Based Permutation : Medical images are tousled by a two-dimensional area preserving map. All pixels in that image is scanned from Upper Left Corner to Lower Right Corner to construct puzzled image. For shuffling the square image some maps used. There is an Arnold cat map, Standard map and Baker map. For shuffling non-square image, Pixel Swapping

based Permutation Strategy(PSP) is urbanized. Here, Pixels of the Plaintext and the Ciphertext are represented commencing Top to bottom and Left to Right. Every pixel in the source image will be altered with other pixel. All pixels are swapping using Pixel Swapping based Permutation Strategy.

Process : First, CS with cryptographic features used to compress the plaintext. This is known as the first level Encryption. Chebyshev map is serving as the secret key. Pixel Swapping based Permutation Strategy is used for image permutation. Quantized dimensions are encrypted through Permutation-Diffusion type Chaotic cipher. This method gives extra security level and quality.

J. Vinay pandey et.al [10] presents an algorithm to defend the communication of Medical Images. This work based on Cryptography, Data Hiding and Steganography. Stream Cipher Algorithm is used for Encrypt the input image. Lossless Data Embedding used to Combine the encrypted image with patient information. Steganography is used to entrench image through the private key. Decrypt the image, inverse methods in reverse order to get the original image and patient information. Extract the image before the decryption of the message to remove the noise in Medical Image.

Stream Cipher : It is also called as Flux Ciphering. It is defined as Block cipher. It has a unitary dimension such as 1 bit, 1 byte. Also, it has a patchy length of 1 to 256 bytes. Stream cipher used for their speeds and ability to modify every sign of the original text. The content of the record is perplexed to the right position and XOR operation is applied.

Process : The original image is enciphered through stream cipher. Then, using the Lossless Data Embedding method, Embed the encrypted image with Patient information. Apply Reversible Data Hiding algorithm to eliminate the entrenched data on enciphering image. Apply Steganography in embedded image among the Private Key. Apply the contrary methods in reverse order to get the original image and Patient details. Apply extraction method to eliminate the noise before Decryption. Existing methods have Less security and more noise. Steganography gives more security. The

Reversible Data hiding method used to remove the noise. This method is Fast and Less Noisy Image retrieved.

V. CONCLUSION

In this Paper, I have surveyed various image encryption and decryption techniques. The Security of images is very important today. Many encryption techniques are studied and analyzed to endorse the recital of the encryption methods. In all methods, Original image is embedded and encrypted then send it to the Receiver. Each Algorithm, Method and Technique used are unique. Every day latest encryption technique is evolving. More secure encryption techniques with high rate of security will work out forever. We should find further secured and a reduced amount of noisy medical image.

VI.REFERENCES

[1] Q. N. Natsheh, B. Li, A. G. Gale, "Security of multi-frame DICOM images using XOR encryption approach", International Conference On Medical Imaging Understanding and Analysis 2016.

[2] Vratesh Kumar Kushwaha, K. Anusudha, "ROI Based Double Encryption Approach for Secure Transaction of Medical Images", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol.2, Issue 4, April 2013.

[3] Simranjeet Kau r , Sukhjinder Kaur, Birdevinder Singh, "Data Hiding Technique for Secure Transmission of Medical Images", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163, Volume 1 Issue 8 (September 2014).

[4] C. Deepak Naidu, Srinivas Koppu, V. Madhu Viswanatham, and S.L Aarthy, "Cryptography Based Medical Image Security with LSB Blowfish Algorithms," ARPN Journal of Engineering and Applied Sciences, VOL. 9, NO. 8, AUGUST 2014.

[5] A.Umamageswari, G.R.Suresh , "A New Cryptographic Digital Signature for Secure Medical Image Communication in Telemedicine," International Journal of Computer Applications (0975 – 8887) Volume 86 – No 11, January 2014.

[6] P. Antony raj, Mrs. A. Umamageswari, "Enhancing Security in Medical Image Communication using Digital Signature," International Journal of Computing Communication and Information System(IJCCIS) Vol 6. No.1 – Jan-March 2014 Pp. 29-34.

[7] Boukhatem Mohammed Belkaid, Lahdir Mourad, "Secure Transfer of Medical Images Using Hybrid Encryption",2015.

[8] Amarit Nambutdee, Surapan Airphaiboon, "Medical Image Encryption based on DCT-DWT Domain Combining 2D-DataMatrix Barcode",2015.

[9] Li-bo Zhang, Zhi-liang Zhu, Ben-qiang Yang, Wen-yuan Liu, Hong-feng Zhu and Ming-yu Zou,"Medical Image Encryption and Compression Scheme using compressive Sensing and Pixel Swapping Based Permutation Approach", Research Article, 13 July 2015.

[10] Vinay Pandey, Angad Singh, Manish Shrivastava, " Medical Image Protection by using Cryptography Data Hiding and Steganography ", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459), Volume 2, Issue 1, January 2012.

[11] P. Vidhya Saraswathi, M. Venkatesulu, " A Block Cipher based on Boolean Matrices using Bit level Operations", Proceedings of International Conference on Computer and Information Science, pp 59-63, 4th - 6th June 2014.

[12] P. Vidhya Saraswathi, M. Venkatesulu, "A Block cipher for Multimedia Encryption using Chaotics Maps for Key Generation", Proceedings of International Conference, " Advances in Information Technology and Mobile Computing (AIM-2013)", published by Elsevier Science and Technology,pp.277-282, 26th - 27th April 2013.

[13] P. Vidhya Saraswathi, M. Venkatesulu, "A Block cipher for Multimedia Content Protection with Random Substitution using Binary Tree Traversal", Journal of Computer Science, Vol.8, No 9,pp. 1541-1546, August 2012.

[14] P. Vidhya Saraswathi, M. Venkatesulu, " A Class of Boolean Matrices Possessing Inverses Under XOR and AND Operations", European Journal of Scientific Research", Vol.118, No.1, pp. 108-112, January 2014.

[15] P. Vidhya Saraswathi, M. Venkatesulu, "A Secure image Content Transmission using Discrete chaotic maps", Jokull Journal, Vol.63, No.9, pp.404-418, September 2013.

[16] P. Karthika, P. Vidhya Saraswathi, "A Survey of Content based Video Copy Detection using Big Data", International Journal of Scientific Research in Science and Technology, ICASCT2401 | ICASCT | March-April-2017.

Survey on Medical Image Compression Using Contourlet Transform

V.Anusuya[1]*, Dr. V. Srinivasa Raghavan[2], S.Divyabharathi [3]
*[1]Computer Science and Engineering, P.S.R. Engineering college, Sivakasi,
pgkrishanu@gmail.com[1]
[2]Principal, S. Veerasamy Chettiar College of Engineering and Technology, Puliangudi,Tamilnadu,India
Vsraghavan1965@yahoo.in[2]
[3]Computer Science and Engineering, P.S.R. Engineering college, Sivakasi, Tamilnadu,India

ABSTRACT

Medical image compression is essential nowadays for telemedicine application. Image compression is associated with removing redundant information and transmits required image data. To transmit large amount of data lot of techniques has been implemented in image compression. Image compression methods already implemented for normal images, one dimensional and 2D image. Recent work has been concentrated on medical images. BPC(BitPlane Coding) technique has been implemented in medical image compression. In this paper, presents a research overview of medical image compression using contourlet transform using wavelet transform, EBCOT algorithm and vector quantization, and its techniques with its future scenario.

Keywords : Medical image Compression, EBCOT, vector quantization, Image processing, Encoding, Wavelet Transform.

I. INTRODUCTION

Digital images require huge amounts of space for storage and large bandwidths for transmission. A 640 x 480 colour image requires close to 1MB of space. Image compression addresses the problem of reducing the amount of data required to represent a digital image. It is a process intended to yield a compact representation of an image, thereby reducing the image storage/transmission requirements.The goal of image compression is to reduce the amount of data required to represent a digital image. Reduce storage requirements and increase transmission rates. Compression is achieved by the removal of one or more of the three basic data redundancies:

    i. Coding Redundancy
    ii. Interpixel Redundancy
    iii. Psychovisual Redundancy

i) Coding Redundancy:
Coding redundancy is present when less than optimal code words are used.

* Code: a list of symbols (letters, numbers, bits etc.)

ii) Inter pixel redundancy:

Interpixel redundancy results from correlations between the pixels of an image.

Interpixel redundancy implies that any pixel value can be reasonably predicted by its neighbours (i.e., correlated).To reduce interpixel redundancy, the data must be transformed in another format (i.e., through a transformation)

E.g Thresholding, differences between adjacent pixels, DFT.

iii) Psycho visual redundancy:
Psycho visual redundancy is due to data that is ignored by the human visual system (i.e. visually non essential information).

* The human eye does not respond with equal sensitivity to all visual information.
* It is more sensitive to the lower frequencies than to the higher frequencies in the visual spectrum.
* Idea: discard data that is perceptually insignificant!

A. Image Compression Model

Image compression techniques reduce the number of bits required to represent an image by taking advantage of these redundancies. An inverse process called decompression (decoding) is applied to the compressed data to get the reconstructed image.

Image compression systems are composed of two distinct structural blocks: an encoder and a decoder.
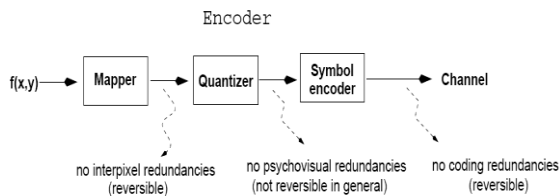
Fig1 Block Diagram for Encoder
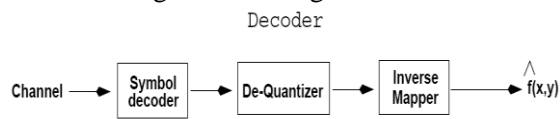
Fig 1.2 Block Diagram for Decoder

B. Image Compression Techniques

The image compression techniques are broadly classified into two categories depending whether or not an exact replica of the original image could be reconstructed using the compressed image. These are:
  i). Lossless technique
  ii). Lossy technique.

i) Lossless Compression

In lossless compression techniques, the original image can be perfectly recovered from the compressed (encoded) image. These are also called noiseless since they do not add noise to the signal (image).It is also known as entropy coding since it use statistics/decomposition techniques to eliminate/minimize redundancy. Lossless compression is used only for a few applications with stringent requirements such as medical imaging. Lossless compression is preferred for artificial images such as technical drawings, text and medical type images, icons or comics. Some of the techniques included in lossless compression are:

  a. Run length encoding
  b. Huffman encoding
  c. LZW coding
  d. Area coding

ii) Lossy Compression

Lossy schemes provide much higher compression ratios than lossless schemes. Lossy schemes are widely used since the quality of the reconstructed images is adequate for most applications. By this scheme, the decompressed image is not identical to the original image, but reasonably close to it. In lossy compression the recovered data is a close replica of the original with minimal loss of data. Lossy compression is used for signals like speech, natural images, etc. Lossy compression techniques include the following schemes:

  a. Transformation coding
  b. Vector quantization
  c. Fractal coding
  d. Block Truncation Coding

Metrics for evaluating compression algorithms

C. Quantitative measures

Two error metrics are often used to compare the original and the reconstructed image following the use of lossy image compression. These metrics are the mean square error (MSE) and the peak signal-to-noise ratio (PSNR).

The MSE and PSNR are routinely published in the literature and are frequently used to quantitatively compare lossy compression techniques. However, MSE and PSNR often do not correlate well with the subjective quality perceived by the human visual system.

D. Qualitative Measures

Receiver Operating Characteristic (ROC) analysis is used to measure the difference in perceived quality between the original and reconstructed image. Observers with expertise in diagnosing disease review a series of images of a disease that have been compressed to a predetermined ratio and then reconstructed. For each image, the observer assigns an ROC confidence rating on a predefined scale representing their impression of the likelihood the disease is present.

These results are compared with the diagnosis made from the original image. If the results are very close for both the original and the reconstructed image, then this can indicate an acceptable level of

compression for this particular disease. A meaningful ROC analysis often requires many images (> 100) and several observers. Although this method is tedious and time-consuming, it is an accepted method in the radiology community for determining image quality based on visual perception.

Image compression may be lossy or lossless. Lossless compression is preferred for archival purposes and often for medical imaging, technical drawings, clip art, or comics. Lossy compression methods, especially when used at low bit rates, introduce compression artifacts. Lossy methods are especially suitable for natural images such as photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate. Lossy compression that produces negligible differences may be called visually lossless.

The organization of this document is as follows. In Section 2 (**LITREATURE SURVEY**), I'll give detail of methods and Techniques used in various papers and also discussed the advantages and disadvantages. In Section 3 (**Future work**), based on the above discussion we have decided the research work. Discussed in Section 4 (**Conclusion**) Comparing the performance of various system.

II. LITERATURE SURVEY

To analyse various medical image compression techniques literature survey has been made.
Santhosh, B., &Viswanath, K. (2016). (pp. 531-537). Springer India.[1] system proposed a algorithm to reduce noise in medical images using contourlet transform. To prove the performance of contourlet transform two types of noises such as Gaussian noise and speckle noise added into an image.

F. Aulí-Llinàs, IEEE Trans. Multimedia, vol. 17, no. 8, pp. 1385–1390, Aug. 2015[2]

The system proposed an algorithm context-based Adaptive Binary Arithmetic Coding (CABAC) to compress video. It uses arithmetic coding and context modeling for providing reduction in redundancy and adaptation. The system proposes two algorithms FLW, FL2W. It simplifies implementation and reduces computational complexity.

F F. Aulí-Llinàs and M. W. Marcellin, IEEE Trans. Multimedia, vol. 16, no. 4, pp. 960–970, Jun. 2014 [3]

The system proposed parallel processing. In the existing system context adaptive arithmetic coding gives poor coding performance in spatial scalability, to enhance the performance it performs bit plane coding and stationary probabilistic model in parallel. The performance of proposed system is similar when the code block size is medium and Large, but it gives better performance if the code block size is small.

F. Aulí-Llinàs, M. W. Marcellin, J. Serra-Sagristà, and J. Bartrina-Rapesta Inf. Sci., vol. 239, no. 1, pp. 266–282, Aug. 2013[4].

The proposed system achieve parallelism implements bit plane coding code blocks given tostatistical probabilistic model to capture the statistical behavior of the image. This algorithm supports all types of 3-D images. An important aspect of this mathematical framework is its generality, which makes the proposed scheme suitable for different types of 3D images. The features of the proposed systems are competitive coding performance, low computational load, very low memory requirements, straightforward implementation, and simple adaptation to most sensors.

Francesc Aulı-Llinas, Pablo Enfedaque et al Proceedings of 2015 Data Compression Conference pp. 163-172 [5]

The system proposed microscopic parallelism it is achieved by multiple coefficient coded in parallel. Biplane coding coefficients are given to arithmetic coding. This system suitable for SIMD and MIMD processor.

P. Eben Sophia and J. Anitha Intelligent Decision Technologies 10 (2016)

The proposed system input is 3-D image. Image split into regions using prediction algorithm. ROI images coded using arithmetic coding and NROI coded using contourlet transform then the values coefficients of the transformed coefficients are normalized using mathematical approach and then the normalized value quantized using arithmetic coding. The performance of the system analysed by using a measure

Compression ratio (CR), Peak signal to noise Ratio(PSNR).Compression ratio of ROI is less compared to NROI. But the PSNR is high for ROI, low for NROI.

NehaMadhukumar, Sanchez V, Abugharb, ijarece,2015[7]

In this system input is a 3-D image given to contourlet transform the K-space is formed by using Fourier transform then 2-D random sampling applied to segment the image. Reconstruction is achieved by applying the reverse process.

Sanchez V, Abugharbieh R, Nasiopoulos P: IEEE Trans Med Imag (7):1062–1072,2009 [8]

A Several techniques based on the three–dimensional (3-D) discrete cosine transform (DCT) have been proposed for volumetric data coding. These techniques fail to provide lossless coding coupled with quality and resolution scalability, which is a significant drawback for medical applications. This paper gives an overview of several state-of-the-art 3-D wavelet coders that do meet these requirements and proposes new compression methods exploiting the quad tree and block-based coding concepts, layered zero-coding principles, and context-based arithmetic coding. Additionally, a new 3-D DCT based coding scheme is designed and used for benchmarking. The proposed wavelet-based coding algorithms produce embedded data streams that can be decoded up to the lossless level and support the desired set of functionality constraints. Moreover, objective and subjective quality evaluation on various medical volumetric datasets shows that the proposed algorithms provide competitive lossy and lossless compression results when compared with the state-of-the-art.

F. Aulí-Llinàs and J. Serra-Sagristà, IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 7, pp. 923–936, Jul. 2008[9].

Input is a 3-D image. In 3-d image transformation performed generates layered bit stream. Scalability is achieved by using rate distortion optimization techniques. The performance is measured by comparing the quality of original image and reconstructed image.

T.-T. Wong, C.-S.Leung, P.-A.Heng, and J. Wang, IEEE Trans. Multimedia, vol. 9, no. 3, pp. 668–673, Apr. 2007.[10]

Image is processed by Discrete wavelet transform (DWT) it leads poor performance when compressing large volume of data. This system proposed to implement convolution based DWT. and this paper used SIMD algorithm to achieve parallel processing in a single processor.

M. N. Do, M. Vetterli, IEEE Transactions on Image Processing, no. 12, pp. 2091-2106, 2005[11]

Wavelet transform is multi resolution in nature. Apply 2-D transform to image it capture geometric structure for better visual information. To capture discrete nature by applying contourlet transform.

Srikanth R, Ramakrishnan AG: IEEETransMedImag24(12):1199–1206, 2005[12]

The system proposed mesh based coding for 3-D medical images.mesh based scheme removes irrelevant background and content based mesh finds optical flow between two spatial edges. The performance of this system is compared with uniform and adaptive mesh-based schemes. The proposed system performance is better than existing system.

III. FUTURE WORK

The emphasis of this paper is on lossless coding of 3-D MRI images, which is a primary requirement of our collaborators. However, the recent 3-D compression schemes for medical images provide important functionalities like region of interest coding and progressive transmission of images. The schemes also provide additional functionality of decoding 2-D images or any objects of interest from the 3-D encoded images. The current implementation of our work does not provide these important functionalities.

A future work the combination of the coefficients from both transforms into zero trees (representing the inter scale dependencies) will be studied. Future work involves compressing other types of medical images with this proposed system using parallel computing

IV. CONCLUSION

Input is an 3-D MRI image (Dicom or Bmp) then process the input image using contourlet transform. The 3D medical images decomposed into 2D slices and then given to Contourlet transform. It is a multi-resolution and directional decomposition of a signal using a combination of Laplacian Pyramid (LP) and a Directional Filter Bank (DFB). The LP decomposes images into sub bands and DFB analyses each detail image. The co efficients of the transform are then quantized using different quantization levels for each subband. Namely, more levels are assigned to important subbands and scales. The encoder takes an input vector and outputs the index of the codeword that offers the lowest distortion. In this case the lowest distortion is found by evaluating the Euclidean distance between the input vector and each codeword in the codebook. Once the closest codeword is found, the index of that codeword is sent through a channel (the channel could be a computer storage, communications channel, and so on). When the decoder receives the index of the codeword, it replaces the index with the associated codeword. Encode the image using vector quantization, then make reverse for decoding, for pre processing the image can use any technique

V. BENEFITS OF COMPRESSION

- Storage Space compressing data files allows one to store more files in the storage space that is available
- Bandwidth and Transfer Speed Compressed files contain fewer "bits" of data than uncompressed files, and, as a consequence, use less bandwidth when we download them.
- Cost of storing the data are reduced by compressing the files for storage because more files can be stored in available storage space when they are compressed. [8]
- Accuracy also reduces the chance of transmission errors since fewer bits are transferred [9].
- Security also provides a level of security against illegitimate monitoring [9].

VI. REFERENCES

[1]. Santhosh, B., &Viswanath, K. (2016). Review on Secured Medical Image Processing. In Information Systems Design and Intelligent Applications (pp. 531-537). Springer India

[2]. F. Aulí-Llinàs, "Context-adaptive binary arithmetic coding with fixed-length codewords," IEEE Trans. Multimedia, vol. 17, no. 8, pp. 1385–1390, Aug. 2015

[3]. F. Aulí-Llinàs and M. W. Marcellin, "Stationary probability model for microscopic parallelism in JPEG2000," IEEE Trans. Multimedia, vol. 16, no. 4, pp. 960–970, Jun. 2014

[4]. F. Aulí-Llinàs, M. W. Marcellin, J. Serra-Sagristà, and J. Bartrina-Rapesta, "Lossy-to-lossless 3D image coding through prior coefficient lookup tables," Inf. Sci., vol. 239, no. 1, pp. 266–282, Aug. 2013.

[5]. FrancescAul´ı-Llin`as, Pablo Enfedaque, Juan C. Moure‡, Ian Blanes†, and Victor Sanchez§ "Strategy of Microscopic Parallelism for Bitplane Image Coding".

[6]. [6] P. Eben Sophia∗ and J. Anitha," Contourlet transform based subband normalization for region based medical image compression" Intelligent Decision Technologies 10 (2016) 385–391.

[7]. NehaMadhukumar, Sanchez V, Abugharb, "Contourlet Transform Based MRI Image Compression using Compressed Sensing" International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 6, June 2015

[8]. Sanchez V, Abugharbieh ,R, Nasiopoulos P: Symmetry-based scalable lossless compression of 3-D medicalimagedata.IEEETransMedImag28(7):1 062–1072, 2009

[9]. F. Aulí-Llinàs and J. Serra-Sagristà, "JPEG2000 quality scalability without quality layers," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 7, pp. 923–936, Jul. 2008.

[10]. T.-T. Wong, C.-S.Leung, P.-A.Heng, and J. Wang, "Discrete wavelet transform on consumer-level graphics hardware," IEEE Trans. Multimedia, vol. 9, no. 3, pp. 668–673, Apr. 2007.

[11]. M. N. Do, M. Vetterli, "The contourlet transform: An efficient directional multi resolution image representation", IEEE Transactions on Image Processing, no. 12, pp. 2091-2106, 2005

[12]. Srikanth R, RamakrishnanAG: Contextual encoding in uniform and adaptive mesh-based lossless compression of MRimages.IEEETransMedImag24(9):1199–1206, 2005

[13]. http://www.ijcset.com/docs/IJCSET15-06-05-020.pdf.

[14]. http://www.ijarcce.com/upload/2015/march-15/IJARCCE%20117.pdf

[15]. http://www.ijcstjournal.org/volume-2/issue-4/IJCST-V2I4P10.pdf

[16]. https://www.tutorialspoint.com/dip/introductio n_to_jpeg_compression.htm

[17]. https://www.google.co.in/ ?gfe_rd=cr&ei=eGnLWN7VMePx8AfL16TY BQ&gws_rd=ssl#q=spacial+redundancy+ppt&

[18]. https://www.google.co.in/?gfe_rd=cr&ei=eGnL WN7VMePx8AfL16TYBQ&gws_rd=ssl#q=sp acial+redundancy+formula&*

[19]. http://www.pcmag.com/encyclopedia/term/4 6335/lossy-compression

[20]. http://en.wikipedia.org/wiki/Vector_quantizatio n rimtengg.com/coit2007/proceedings/pdfs/ 43.pdf

An Efficient Load Balancing Scheme in Cloud Computing

D. Pradhiba*[1], R.Kanniga Devi[2]
*[1]P.G Student, Department of Computer Science and Engineering, Kalasalingam University, Srivilliputtur,Tamilnadu, India[1]
[2]Assistant Professor, Department of Computer Science and Engineering, Kalasalingam University,Srivilliputtur,Tamilnadu,India[2]

ABSTRACT

In computer science domain, cloud computing is one of the booming technologies. In cloud computing platform the clients and cloud data centers are distributed across the world. The biggest challenge of cloud data centers is servicing and handling millions of request from the clients. Here the major challenge is load balancing the cloud datacenter. Load balancing is a technique which is used to spread the load of system fairly across all the servers and avoiding overloading. The prime goal of balancing the load of system is to minimize fail-over, attain better resource utilization and enabling scalability etc. In this paper, an algorithm is proposed to distribute load of the system uniformly across cloud data center for balancing the load which decreases average execution time of cloud data centers.

Keywords: Cloud computing, Cloud Datacenter, Load balancing

I. INTRODUCTION

The cloud computing is one of the most trending technologies in IT domain. It is a technique of handling and pooling services like servers, data base, storage, software and more over the internet based on the user's need or demand. Users can get the resources from the data centers as per their requirements from anywhere through an internet connected computer or hand held devices. One of the challenging task in cloud computing is load balancing used to allocate work load among the data centers. Datacenters are physical machines that has the responsibility to complete the request and demand of cloud users. So load balancing is required to manage the load across data centers, reduce the overload, improve performance, minimize average execution time and provide better resource utilization. Load balancing can minimize the response time and maximize the user's satisfaction. It also increase the source utilization and limit the energy consumption. The classification of load balancing algorithms are of two types: static and dynamic. Load balancing is static when it needs previous data of system. It is dynamic when it requires current data of system.

In static load balancing algorithm [1], load is assign to machine as per their capacity of processing. It has preliminary knowledge of the system like performance, processing energy, memory and data about client requirement. It doesn't consider the dynamic changes at run time. Popular static algorithms like Round Robin algorithm, Weighted Round Robin algorithm are used for balancing the load.

In dynamic load balancing algorithm,[1] preliminary information about the system is not required, because runtime condition of system is collected for assign load to every machine. Popular dynamic algorithms like Least Connection algorithm, Weighted Least Connection algorithm are used for balancing the load. The dynamic algorithm gives higher performance and fault tolerance, but it is difficult.

Dynamic load balancing helps to improve the efficacy, maintaining the stability of the system and adjusting system modification. In this paper, an algorithm is proposed to distribute load of the system uniformly

across cloud data center for balancing the load which decreases average execution time of cloud data centers.

II. RELATED WORK

An Gulshan soni, et al.[1] proposed the central load balancer technique that has the responsible to take decision for balancing the workload among virtual machines. Central Load Balancer (CLB) have VM table which contain VM state, priority of VM .The allocation of request to VM is based on the state and priority of the VM. It tried to avoid the situation of under loading and over loading of virtual machines.

Fei Ma ,et al.[2] used TOPSIS method for allocation of virtual machine in cloud data centers to achieve load balancing. In this system, two decision have to be taken for balancing the load when the system is over loaded or under loaded. The VM monitor observes the over and under loaded of the system. The PM have to be migrated to selected VM and which VM have to be migrated when the PM is problematic. In large scale of cloud environment this system achieve better load balancing with less migration time.

Vasudha Arora, et al.[3] analyzed the performance of round robin algorithm, equally spread execution load and throttled load balancing policies across virtual machine. The throttled load balancing policy reduces average response time and data processing time of data centers. But it's not suitable when data centers are reconfigured for route the traffic dynamically.

Feilong tang et al.[4] proposed a novel dynamical load-balanced scheduling (DLBS) approach to maximize the network throughput by balancing workload dynamically. The results indicate that this DLBS approach is able to balance transmission traffic dynamically and globally so that it improves throughput and bandwidth utilization ratio.

Surbhi kapoor et al.[5] The problems of existing throttled and modified throttled algorithms are not considering the resource specific demands of the tasks and also they are not suitable for heterogeneous VMs environment. These problems have been addressed in this proposed approach by clustering the VMs.

Youssef fahim, et al.[6] proposed the hybrid algorithm to improve the availability, performance and maximize the VM utility. This hybrid algorithm takes into account the current load of VM and current task status of VM during execution to avoid probable blocking.

Ekta gupta, et al.[7] proposed a load balancing technique which is based on Ant Colony Optimization .The under loaded and over loaded servers are identified and load balancing operation are performed by Ant Colony Optimization between identified datacenter server to increase the throughput and availability of resource.

Joseph doyle, et al.[8] considered carbon emission, average service request time and electricity cost as balancing factor. The voronoi partitions method was used to identified which request is allocated to which data centers as per balancing factors in order to balance the workload of system.

Chung-Cheng Li, et al.[9] proposed a novel neural network-based dynamic weighted round-robin scheduling algorithm and SLA-aware two-level decentralized load balancer (tldlb) architecture to achieve dynamic load balancing in cloud datacenter. To distribute the large number of incoming request to different VM a novel neural network-based dynamic load balancing algorithm was developed and SLA violation rate was reduced by using a novel decentralized load balancing architecture.

G.Shobana, et al.[10] proposed preemptive task scheduling which follows the foraging behavior of honey bees to achieve load balancing in cloud datacenter. Tasks are allocated to VMs based on the foraging behavior of honey by preemptive task scheduling which reduces the make span. It also considered task priority to minimize latency and maximize throughput.

Hitesh A. Bheda, et al.[11] used the flexible load sharing algorithm which partitioned the system into domain for load balancing in cloud computing. It overcomes the message loss problem and it provides unlimited resource capacity. Biased random selection method was used by this FLS algorithm for the scalability and performance improvement.

Ankit Kumar et al.[12] proposed a modified active monitoring load balancer which distributes the incoming jobs among the VM efficiently. Whenever a request comes from client it has sent to modified active monitoring load balancer which have VM table. Then it checks the VM table for allocating the requests to VM. If VM is available it allocates the request to the VM. Otherwise it waits until the VM became available state. The VM table contains VM ids, state of VMs, VM loads and memory of VMs. This algorithm does not consider the VM reliability and energy awareness.

J.Octavio, et al.[13] used the distributed problem solving techniques for balancing the heterogeneous loads optimally. The CProtocol and EProtocol was used by agent for manage the load in distributed manner, it reduce the cost of energy consumption.

III.PROPOSED WORK

In this section, we will be discussing our proposed approach. In this proposed approach, cluster heads were elected from group of data centers depends upon the higher number of connection with other data centers. The data centers were clustered based on the threshold value or capacity which is assigned by the cloud service provider. The input of each client is assigned to cluster head based on the shortest path between client and cluster head. In previous load balancing schemes there is a variation in load of each data centers (some data center may have high load and some have low load) because of shortest path. To overcome this our proposed approach assigns capacity to each data center. If the capacity of data center exceeds then the remaining client job has transferred to the nearest data center which has a capacity to serve. The cluster heads were communicated with other to know the status and count of services provided. By this new approach data centers can achieve uniform load balancing.

This proposed scheme is implemented in c platform by using graph theory concept. The step-wise procedure of the proposed approach is given as follows:

Steps to achieve uniform load balancing:

Clustering data centers:

Step 1: Datacenter Network is consider as a undirected weighted graph G= (V, E, W) .Where V is datacenter, E is connection between datacenters and W is weight age of the connection between datacenters.

Step 2: In order to identified the cluster head (CH) the degree value should be calculated for each datacenter (V)

Step 3: The datacenters which are having maximum degree value have to be considered as cluster heads (CH).

Step 4: Then the remaining datacenters (V') have to be allocated to cluster heads (CH) based on the shortest path distance, the shortest path distance is calculated by weight age of the connection between datacenter (V') and cluster heads.

Step 5: The limit of datacenters (V') which are assigned to the cluster heads(CH) are based on the threshold value.

$$\text{Threshold} = \frac{\text{Number of datacenters (V)}}{\text{Number of clusters (K)}}$$

Where the number of clusters (K) should be defined by cloud provider

Client assignment to clusters:

Step 6: Each client has to be allocated at most uniformly to the clustered datacenters (CL) based on the threshold value.

$$\text{Threshold} = \frac{\text{Number of clients(C)}}{\text{Number of clusters (k)}}$$
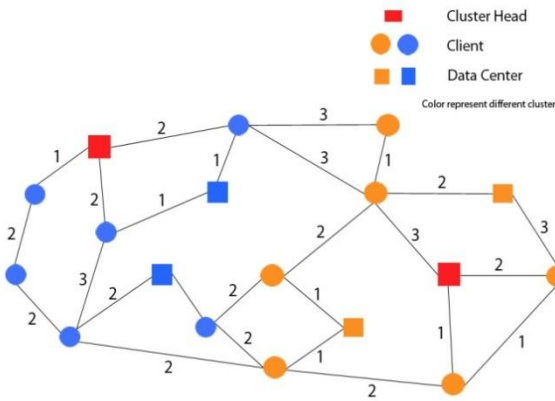
Figure 1: Model of proposed system

Figure 2 Comparison of existing and proposed algorithm

IV. EXPERIMENT AND RESULTS

The experiment has been carried out by implementing the proposed algorithm and the existing algorithm in C. A random task generator have been used. There is no limit to the number of data centers generated. Implementation parameters have been mentioned in table 1. Our proposed algorithm has been compared with K-Nearest neighbor algorithm and shared nearest neighbor clustering algorithm on the basis of execution time and response time.

Table 1 Implementation Parameters

Number of Data centers	9-100
Number of Clusters	3
Number of Clients	18-300

Comparison of proposed and other two algorithms have been shown in table 2. By varying the data centers and clients the comparison table is shown for the three algorithms on the basis of execution time and response time.

Table 2 Comparison Table

Clients	Neighbor size K-value	Cluster Time (sec)		
		Proposed algorithm	K-Nearest Neighbor algorithm	Shared Nearest Neighbor algorithm
69	23	184	212	201
135	45	320	351	335
165	55	452	485	471
196	32	500	528	515
291	97	579	616	599

Figure 2 shows the comparison of existing algorithms with proposed algorithm. It clearly shows that the clustering time is lesser than the previous algorithms.

V.CONCLUSION

The biggest challenge in cloud computing is balancing the load of cloud datacenters. So load balancing is required to manage the load across data centers to avoid overload of datacenter, improve performance of datacenters, minimize average execution time and provide better resource utilization. we present a load balancing algorithm to distribute load of the system uniformly across cloud data center for balancing the load and decreases average execution time of cloud data centers.

VI.REFERENCES

[1] Gulshan Soni, Mala Kalra,"A Novel Approach for Load Balancing in Cloud Data Center", IEEE International Conference on Advance Computing, pp. 807-812,2014.

[2] Fei Ma, Feng L1U, Zhen L1U, "Distributed Load Balancing Allocation Of Virtual Machine In Cloud Data Center", IEEE International Conference On Computer Science And Automation Engineering, pp. 20– 23.

[3] Vasudha Arora, S. S.Tyagi, "Performance Evaluation of Load Balancing Policies across Virtual Machines in a Data Center", International Conference on Reliability, Optimization and Information Technology,pp.384-387,2014.

[4] Feilong Tang, Laurence T. Yang, Can Tang, Jie Li, Minyi Guo, "A Dynamical And Load-

Balanced Flow Scheduling Approach For Big Data Centers in Clouds", IEEE Transactions On Cloud Computing, Volume: 1, Issue: 99, 2016.

[5] Surbhi Kapoor, Dr. Chetna Dabas, "Cluster Based Load Balancing In Cloud Computing", IEEE Eighth International Conference On Contemporary Computing, pp. 76 - 81, 2015.

[6] Youssef Fahim, Elhabib Ben Lahmar, El Houssine Labriji Saea Ouahabi, "The Load Balancing Improvement of A Data Center By A Hybrid Algorithm in Cloud Computing", IEEE Third International Colloquium In Information Science And Technology,pp.141-144,October 2014

[7] Ekta Gupta, Vidya Deshpande, "A Technique Based On Ant Colony Optimization for Load Balancing in Cloud Data Center", IEEE 13th International Conference On Information Technology, pp. 12-16, 2014 .

[8] Joseph Doyle, Robert Shorten, Donal O'Mahony, "Stratus: Load Balancing the Cloud for Carbon Emissions Control", IEEE Transactions On Cloud Computing, Vol. 1, No. 1,pp. 2168-7161, June 2013.

[9] Chung-Cheng Li, Kuochen Wang, "An SLA-Aware Load Balancing Scheme for Cloud Data Centers", IEEE International Conference On Information Networking, pp. 58 - 63,2014

[10] G.Shobana, M.Geetha, .R.C.Suganthe, ",Nature Inspired Preemptive Task Scheduling for Load Balancing in Cloud Datacenter", IEEE Information Communication and Embedded Systems, 2014 International Conference on , pp. 1-6, 2014.

[11] Hiren H. Bhatt, Hitesh A. Bheda, "Enhance Load Balancing Using Flexible Load Sharing In Cloud Computing", IEEE 1st International Conference On Next Generation Computing Technologies, pp.72-76, September 2015.

[12] Ankit Kumar, Mala Kalra, "Load Balancing in Cloud Data Center Using Modified Active Monitoring Load Balancer", IEEE International Conference on advances in Computing, Communication and Automation, pp. 1-5, Apr 2016.2016

[13] J.Octavio, Gutierrez-Garcia, Adrian, Ramirez-Nafarrate, "Collaborative Agents for

Distributed Load Management in Cloud Data Centers Using Live Migration of Virtual Machines", IEEE Transactions On Services Computing, Vol. 8, No. 6,pp. 1939-1374, November/December 2015.

[14] Wang Yong, Tao Xiaoling, He Qian, Kuang Yuwen, "A Dynamic Load Balancing Method Of Cloud-Center Based On SDN ",IEEE Transactions On Network Technology And Application, Volume: 13, Issue: 2, pp. 130 – 137, February 2016 .

[15] Jia Zhao, Kun Yang, Xiaohui Wei, Yan Ding, Liang Hu, "A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment," IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 2, pp. 1045-9219, February2016

[16] Lei Yu, Liuhua Chen, Zhipeng Cai, Haiying Shen, Yi Liang, Yi Pan, "Stochastic Load Balancing for Virtual Resource Management In Datacenters", IEEE Transactions On Cloud Computing, Vol. , No. , November 2014.

A Survey of Heart Disease Prediction Using Classification Techniques

C. Sowmiya[1], Dr. P. Sumitra[2]
[1]Ph.D Research Scholar, Department of Computer Science,
[2]Professor, Department of Computer Science,
[1,2]Vivekananda College of Arts and Sciences for Women (Autonomous), Elayampalayam.
sowmiyac83@gmail.com

ABSTRACT

Heart disease is a number one problem for in the world. Every year more than people death for heart disease. This disease for attack not only India all country people affected by heart disease. Some of the people occur deaths for during the first heart attack. Now day's computer field is very high and particular work is finished. So used for medicine area diagnosis. Every year Researchers have been applied data mining techniques and algorithms for diagnosing heart disease. Heart is a very important part of each and every people. Over the past few years more than people deaths, reason from disease is easily attack for patients. Not proper treatment for many hospitals. Mostly affected heart attack for men's. But Women are rare for affected heart attack.

Keywords: Data Mining, Heart Disease, Neural Network, K-Nearest Neighbor Algorithms, Support Vector Machine (SVM), Naive Bayes Algorithm, Decision Tree Method.

I. INTRODUCTION

Now' a day's data mining is very useful for some places and more then details. All the applications of using in world level. Data mining is huge amount of several data base. Data mining is an intelligent creative process. Data mining is used various fields. Now days data mining many places using. This data mining techniques are many advantages and efficient that can be heart disease prediction. Different types of data mining techniques available are classification, cluster, feature selection, association rule can be analyzing the heart disease prediction. . Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific, medical research. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years due to the wide availability of huge amounts of dataData mining is also called as a KDD Process.KDD process means Knowledge Discovery Process.

Now day's data mining using of more than important application fields.

Medical, Insurance, Telecommunication, Finance, Utilities, Data service providers, Transport, Consumer good , IT , Railway, ship, banking.

II. HEART DISEASE

Now day's heart disease is a main reason for death in the world. Heart is very useful part of our human body. This disease is number one problem for in the world. But not only heart attack some of the disease attacked form valve, ventricle, lung cancer, breast cancer ,diabetes etc…..A common heart disease is nothing, but coronary heart disease or cardiovascular disease is a very dangers disease. Most people attack for heart disease from the world. Coronary heart disease blood vessels around it. This disease causes disability as damage to brain so resulting in death. That ranges of the age group 25 to 69 have 25% risk of having heart disease. Some types of

disease occur from heart. There are several kinds of causes, reason, factor which increase the risk factor of heart disease. There are considered as important reason of heart disease. Most of hospitals admitted in heart disease patient. This disease mostly affected in male because smoking habits. This paper analyzes the different kinds of heart disease using the classification techniques.

Types of disease considered are:

Coronary heart disease, angina pectoris, congestive heart failure, cardiomypathy, congenital heart disease, arrhythmias, myocarditis, heart attack; heart cancer etc.

There are considered some important reasons of heart disease:

Age, Smoking, Sugar, Obesity, Depression, Hyper tension, High blood cholesterol, poor diet, Family history ,Physical inactivity.

III. DATA MINING TECHNIQUES

1. Classification
2. Cluster
3. Association rule

Classification

Classification is one of the well known problems under data mining. To classify the data/objects into different classes or groups.   For example data can be broken down according to topical content file type, average file size, gigabytes, and megabytes. Classification is the process of learning a function that cans data objects to a subset of a given class set.   Some types goals of classification, first finding a good general that can predict the class of but far unknown data objects with high accuracy. Second to find a compact and easy Understandable class model for each other classes.

Cluster

Cluster is a group of objects. For example data elements into different groups of similarity between in a single group cluster partitions the data set in to cluster classes. Each and every near object is neighborhood object. There are two goals of cluster. First one is a inter class

second is an intra class. Inter class cluster means cluster distance is maximized. Intra cluster means cluster distances are minimized.

Association rule

Association rule mining is a very import rule of data mining techniques. Association rule is identifying of association huge data base and their values. In this pattern creative techniques which does not serve to solve classification problems and predict problems. This paper focus on study of exiting heart disease prediction task by a using data mining techniques and different issues in exiting of heart disease prediction.

IV. LITERATURE REVIEW

k.srinivas etal(2011) "presented application of data mining techniques in healthcare prediction of heart attacks". The powerful use of classification various data mining techniques using such as a decision tree,k-nearest neighgor naïve bayes. Huge volume of health care data using in the data mining tool Tanagra used for conducted data analysis for the learning purpose. This paper consider of 3000 instances for training data set it is various type 14 attributes. The instances data set is providing the result of various types of testing to say the accuracy of heart disease. The performance of the classifiers set of value and results are analyzed. So to the attributes dataset is divided into two parts. Using training data set is 70% and testing dataset is 30%. The comparison of some classification algorithms using. This paper considered the best algorithm naive bayes.this model using the naive bayes taken a time to run data for best result achieved. comparision of various classifications algorithm hence better performance accuracy in naive bayes.

Algorithm	Accuracy	Time Taken
Naïve bayes	52.33%	609 ms
Decision tree	52%	719 ms
k-nearest neighbor	45.67%	1000 ms

A1-Milli N, (2013) "Back Propagation algorithm method. To improve the multilayer neural networks in a supervised manner. Back propagation algorithm based on error correction learning. This algorithm based through the various layer of network. From the forward and backward class. Compare to oher algorithm better result for back propagation algorithm .this algorithm research to compare classification techniques. The author efficiency and deliver high accuracy from the heart disease prediction.

Anuradha, srinivasaraghavan, vincyjoseph (2016) "Comparative analysis of accuracy on heart disease prediction using classification method "This paper research method algorithm is naïve bayes and SVM (Support Vector Machine), logistic regression.

Naive bayes : Naïve bayes algorithm is a good tool in medical diagnosis. The classifier processes each attributes probability in a class. Naïve bayes is a simple good and efficient performance in classification due to good accuracy for used in medical diagnosis.

Support Vector Machine: SVM is a one of the classification method in this type used to recognize patters and data in a regression and classification analysis. SVM is two classes classified recognizes and separates. Similar data by finding the best hyper plane that separates all data points of one class from other class. Mathematical functions are involved in SVM. Its performance magnify with number of attributes.

Logistic Regression: Logistic Regression is mainly used for prediction besides it can also be used calculating the probability of success. Basically logistic regression involves fitting and equation. It is a type of statistical regression analysis method used for approximation and prediction of result of a dependent attributes. Dependent means it can take only some set of values for example good, or bad, true, or false, on, or off.

These papers consider various classification algorithm using. But high accuracy of performance of heart disease prediction in SVM algorithm is better result.

Algorithm	Accuracy
Naïve bayes	75%
SVM(Support vector machine)	80%
Logistic Regression	79%

Mai Shouman, Tim Turner, Rob Stocker(2011) "Using Decision Tree for Diagnosing Heart Disease Patients". Particular researches have been investigating the application of the various classification techniques. To help health care management with improved accuracy in the diagnosis of heart disease. Andreev a used c4.5 decision tree in the diagnosis of heart disease. Accuracy of 75.73% (Andreev 2006) this paper comparison of 3 classification algorithm c4.5 and j4.8 and bagging algorithm. Considered the best performance algorithm bagging algorithm. When using the UCI repository heart disease data set.

Performance comparison algorithm

Algorithms	Accuracy
J4.8 Decision tree	78.9%
Bagging Algorithm	81.41%
C4.5	75.73%

Neural Network

Neural Network are basically used in biological inspired large inter connected cells that simulate the human brain. The perception is the very simplest architecture. This has one neuron and learning method .many sophisticated architecture is mult-layer neural network. Which are one (or) more

neurons connected various layers. Neural network is based on the two types of training method.1.supervised training method; another method is a 2.un supervised training method.

Decision Tree

Decision tree is a like tree structure format. Decision trees are powerful classification algorithms. It's include the below element.

1. Root node
 2. Branch Node
 3. Leaf node
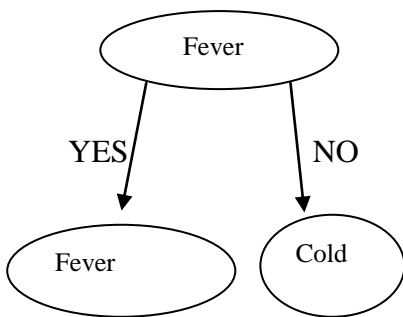
Decision tree advantages:

1. Easy to understand and interpret.
2. It not required the domain knowledge to learn

V. FEATURE WORK

A. Naive Bayesian Algorithm

1. Start the process
2. Load the input user value
3. Initial redail velocity for blood prushur e level (Ex; 100 Pations).
3.1Identify the disease occurrence chance with relative input parameters
3.2Check the Train and test each layer under Neuro- functions
4. Assign a separate weight age to each layer
5. Check the deduction data value (normal & abnormal).
6. Print the result
7. Stop the process....

B. PROGRAM CODE

```
Clear;close all;clc;

load 100_ECG_0_20
s1=ECG_1;s2=smooth(s1,150);ecgsmooth=s1-s2;

[C,L]=wavedec(ecgsmooth,8,'db4');
[d1,d2,d3,d4,d5,d6,d7,d8]=detcoef(C,L,[1,2,3,4,5,6,7,8]);

 [thr,sorh,keepapp]=ddencmp('den','wv',ecgsmooth);
cleanecg=wdencmp('gbl',C,L,'db4',8,thr,sorh,keepapp);
max_value=max(cleanecg);
mean_value=mean(cleanecg);
threshold=(max_value-mean_value)/2;

a5=appcoef(C,L,'db4',5);
C1=[a5;d5;d4;d3];
L1=[length(a5);length(d5);length(d4);length(d3);length(cleanecg)];
R_detect_signal=waverec(C1,L1,'db4');
R_detect_squared=R_detect_signal.^2;

for a=1:length(R_detect_squared)
   if R_detect_squared(a)>threshold
      R_detect_new(a)=R_detect_squared(a);

   else

      R_detect_new(a)=0;

   end

end

mean_R_detect=5*mean(R_detect_new);

for q=1:length( R_detect_new)-1
   if  R_detect_new(q)< mean_R_detect
      R_detect_new(q)=0;

   end

end

d=0;

for b=1:length( R_detect_new)-1
    if ( R_detect_new(b)==0) & ( R_detect_new(b+1)~=0)
    d=d+1;
    indext(d)= b+1;

    end
end
```

```
fs_R_deetect=length(R_detect_new)/20;
time=indext.*1/fs_R_deetect;
ind=0;

for z=1:length(time)-1
    ind=ind+1;
    time_diff(ind)=time(z+1)-time(z);

end

av_time=mean(time_diff);
Square_Number=av_time/.2;
beat_Rate=300/Square_Number;
high=max(R_detect_new);

subplot(411);plot(s1);title('Orginal Signal');
subplot(412);plot(s1-s2);title('Baseline drift
Elimination');
subplot(413);plot(cleanecg);title('Main Signal');
subplot(414);plot(R_detect_new);title('R detected
Signal');
text(length(R_detect_new)/2,high,['BeatRate=',num2str(
fix(beat_Rate))],'EdgeColor','red'
```

C. FLOWCHART
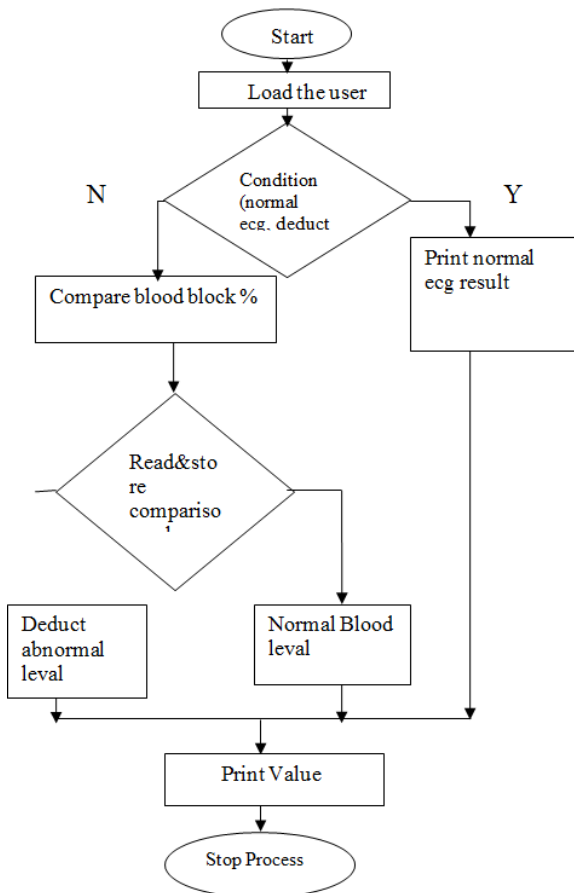
VI. CONCLUSION

In the world every day heart disease problem in attack from people. But every year more than people for death. But not good accuracy presentation in comparisons of various classification algorithm and cluster feature selection method using. This paper considered result. So High accuracy provided from medical field every year death people decrease. The result of the various classification algorithm compared to the Bagging algorithm are very encouraging. The difference in the accuracy is noticeable. This paper the high accuracy and best performance of using Bagging Algorithm 81.41% use the heart disease prediction. It has proven that classification based techniques proposed algorithm – Naïve Bayies contribute high effectiveness and obtain high accuracy compare than the previous methods.Proposed work it can experiment for algorithm Naïve Bayies algorithm

VI.REFERENCES

[1] Han.j.and m.kamber (2006) Data Mining concepts and techniques, Morgan Kaufmann publishers.

[2] K .Viswanathan 1, Dr. K .Mayilvahanan 2 , R. Christy Pushpaleela(2016) Prediction in data minng comparative study of heart disease Elysium journal Engineering Research and Management. Volume-3, Issue-4.

[3] Detrano, R.; Steinbrunn, W.; Pfisterer, M., "International application of a new probability algorithm for the diagnosis of coronary artery disease". American Journal of Cardiology,Vol. 64, No. 3, 1987, pp. 304-310.

[4] Anuradhasrinivasaraghavan,vincyjoseh(2016) "Comparative analysis of accuracy on heart disease prediction using classification method international Journal of applied information systems volume 11-No 2. July 2016.

[5] Chau, M.; Shin, D., "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms". Proceedings of IEEE International Conference on Dependable, Autonomic and Secure Computing 2009, pp.183-187.

[6] Palaniappan, S.; Awang, R., "Intelligent Heart Disease Prediction System Using Data Mining Techniques". Proceedings of IEEE/ACS International Conference on Computer Systems and Applications 2008, pp. 108-115.

[7] Keyue Ding and Kent R Bailey et al, " Genotype-informed estimation of risk of coronary heart disease based on genome-wideassociation data linked to the electronic medical record", International journal of BMC cardiovascular Disorders, Vol.11,2011

[8] Daisy JA Janssen and Emiel FM Wouterset al, "Self-perceived symptoms and care needs of patients with severe to very severechronic obstructive pulmonary disease, congestive heart failure or chronic renal failure and its consequences for their closest relatives:the research protocol", Journal of BMC palliative care, Vol.7, 2008

[9] Shou-En Lu and Gloria L Beckleset al, "Evaluation of risk equations for prediction of short-term coronary heart disease events inpatients with long-standing type 2 diabetes: the Translating Research into Action for Diabetes", International Journal of BMCEndocrine Disorders, Vol.12, 2012.

[10] EmanAbuKhousa, Piers Campbell," Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems", IEEE, International Conference on Innovations in Information Technology, pp.267-272, 2012.

[11] Lamia AbedNoorMuhammed," Using Data Mining technique to diagnosis heart disease", IEEE, International conference on statistics in science, Buiseness and Engineering, pp.1-3, 2012.

Performance analysis of optimally tuned 2DOF-PID controller for Automatic Load Frequency Control

Nivedita G R[*1], Kamaraj N[2]

[*1]Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
gr.nivedita@gmail.com[1]
[2]Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
nkeee@tce.edu[2]

ABSTRACT

In a Power system, if the load varies the frequency of the generator also varies. Automatic load frequency control plays a major role in maintaining the frequency constant. Various controllers like PI, PID and 2DOF-PID are used for controlling the load frequency in power system. Two degree of Freedom Controller optimized by Particle Swarm Optimization will help to adjust the generator frequency to ensure stable and efficient response due to sudden changes in load. The single area power system consisting of non-reheat thermal power plants with 2DOF-PID controller has been considered for design and analysis. The gains of the 2DOF-PIDcontroller is optimized by the particle swarm optimization to have a better dynamic performance, to reject the load disturbance , to improve the robustness and also to reduce the Integral Time Absolute Error which is caused due to sudden changes in load demand.

Keywords: Automatic load frequency control; Transient stability; Two Degree of freedom; Particle Swarm Optimization; Dynamic performance; parameter uncertainties; performance index; steady state error; Integral time absolute error

Abbreviations and Acronyms: K_t-turbine gain; T_t-Turbine time constant; K_g-Governor gain; T_g- Governor time constant; K_{ps}-power system gain; T_{ps}-Power system time constant; R-Governor speed regulation; B-frequency Bias parameter T1-synchronising coefficent;K_p-Proportional gain; K_i-Integral gain; K_d- Derivative gain; ITAE-Integral Time Absolute Error;2DOF -two degree of freedom

I. INTRODUCTION

In an interconnected electrical power system both the voltage and frequency to be fixed at desired values irrespective of change in loads that occurs randomly. To cancel the effect of load variation and to keep the frequency constant a control system is required. Though the active and reactive powers have a combined effect on the frequency and voltage, the control problem of the frequency and voltage can be separated. Frequency is mostly dependent on the active power. The active power and frequency control is called as load frequency control (LFC). Themost important task of LFC is to maintain the frequency constant against the varying active power loads. The main purpose of LFC system are to keep in zero steady state error in frequency deviation for single area system, optimal transient performance and to have reduced oscillation in the system due to frequent change in load demand. The gains of the controller can be optimized by the modern heuristic algorithm (PSO-Particle Swarm Optimization) and also to reduce theperformance index Integral Time Absolute Error (ITAE).

II. AUTOMATIC LOAD FREQUENCY CONTROLLER

A. ALFC system

Various components of Automatic load frequency control loop are shown in Fig.1.
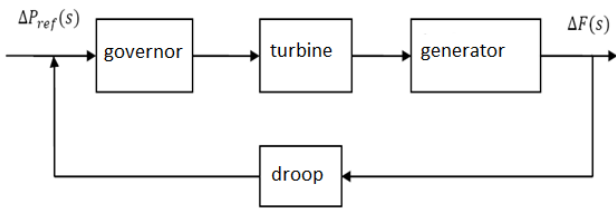
Figure 1 : ALFC   System

A single area ALFC tests system of thermal plants consisting of governor, turbine, and generator. Each component in the ALFC loop is modeled by a first order transfer function system defined by its gain and time constant.

Parameters used in the ALFC system are:

$K_g=1$, $T_g=0.08$, $K_t=1$, $T_t=0.3$, $K_{ps}=120$, $T_{ps}=20$, $R=2.4$,

$T_1=0.545$, $B=0.425$

Governor $G_g=K_g/T_g s+1=1/0.08s+1$.

Turbine $G_t= K_t/T_t s+1=1/0.3s+1$.

Power System $G_{ps}=K_{ps}/T_{ps}s+1=120/20s+1$.

The ALFC single area system model with droop characteristics R can be expressed as an overall transfer function G(s) as:

$$G(s) = \frac{G_g G_t G_{ps} K_g K_t Kps}{1+G_g G_t G_{ps}/R} = \frac{(1)}{(T_g s+1)(T_t s+1)(T_{ps}s+1)+K_{ps}/R}$$

$$G(s) = \frac{G_g G_t G_{ps}}{1+G_g G_t G_{ps}/R} = \frac{120(2)}{(0.08s+1)(0.3s+1)(20s+1)+50}$$

B. ALFC single area system with controllers

The ALFC response can be obtained by using a controller in the forward path. The controller used to analyze the single area test system is PI, PID and 2DOF-PID controllers. The area control error ACE is defined by the equation

ACE=BΔf+ΔPtie(3)
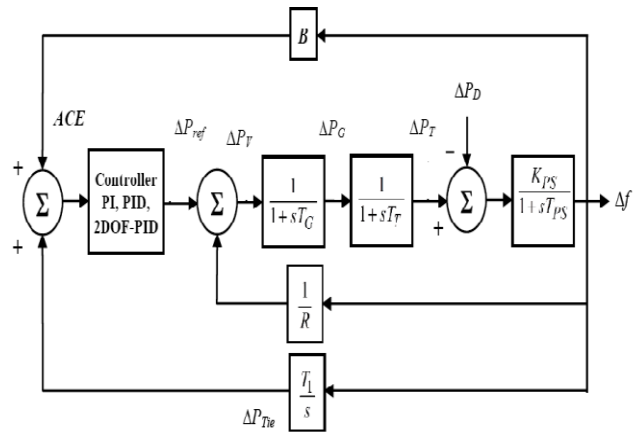
Figure 2 : Single Area ALFC test system with controller
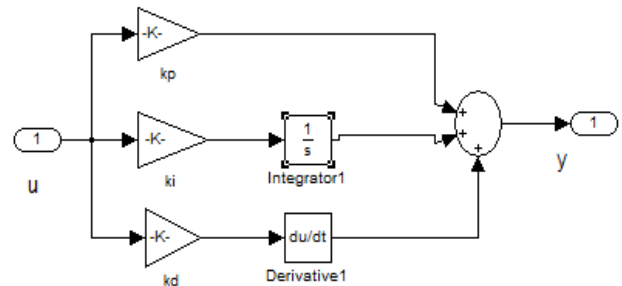
Figure 3: Structure of PID controller

The goal of any controller in a LFC system is that when a disturbance occurs the controller must control the frequency of the system with zero steady state error and less settling time. PI and PID controller used to damp system oscillations, increase stability and reduce steady state error. PID controller continuously calculates an error value as the difference between a desired set point and measured process variable.

Kp, Ki are the gains of proportional, integral part of the controller. These two gains are tuned using PSO.

$$Gc(s) = Kp + \frac{Ki}{s} \quad (4)$$

Kp, Ki, Kd are the gains of proportional, integral, derivative part of the controller. These three gains are tuned using PSO.

$$Gc(s) = Kp + \frac{Ki}{s} + Kds \quad (5)$$

ALFC system with 2DOF PID controller is shown in Fig.5. Two-degree-of-freedom (2-DOF) PID controllers include set point weighting on the proportional term. Two degree of freedom achieves high performance in tracking of set-point and regulation in the presence of

disturbance inputs. The degree of freedom in control system indicates the number of closed loop transfer function that can be adjusted independently. In two degree of freedom controller there is a additional feed forward term which makes a different from the conventional PID controller. Considering that the main advantage of PID controller lies in its simplicity in its structure only the proportional components are added in the feed forward structure.

It consists of Kp,Ki,Kd as in conventional PID and two additional parameters b, c are set points used in the feed forward structure therefore this 2DOFPID is also called as set point weighted PID shown in Fig.4. These set points b and c are assumed as less than one. Due to this feed forward structure of the controller the disturbance due to load change is eliminated and also the controller is robust in case of parameter uncertainties. The load change is given as a step load disturbance. The structure of 2DOFPID is shown in Fig.4
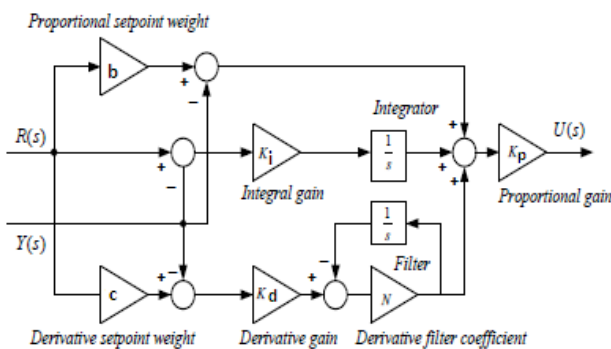
Figure 4: Structure of 2DOFPID controller
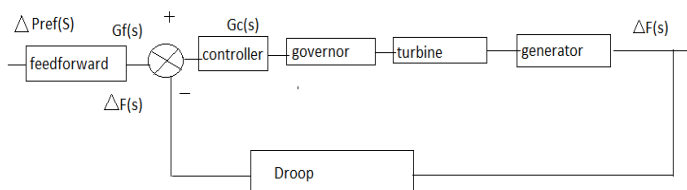
Figure 5: ALFC system with 2DOF-PID controller

The ALFC system with 2DOF-PID controller is shown in Fig 5.The equation $G_c(s)$ and $G_f(s)$ are given by:

$$G_f(s) = \frac{(bK_p+cK_ds^2)+(bK_pN+K_i)s+K_i(6)}{(K_p+K_dN)s^2+(K_pN+K_i)s+KiN}$$

$$G_c(s) = \frac{(K_p+K_dN)s^2+(K_pN+K_i)s+K_iN(7)}{s(s+N)}$$

III. OPTIMIZATION USING PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) is used to explore the search space of a problem to find the parameters required to minimize a particular objective. It is a member of wide category of Swarm Intelligence methods for solving the optimization problems The idea of swarm intelligence based on the observation of swarming habits by animals (such as birds and fish) and the field of evolutionary computation. PSO belongs to the broad class of stochastic optimization algorithms and population-based algorithm that exploits a population of individuals in the search space. The population is called a swarmand the individuals are called particles. Each particle moves with a velocity within the search space, and retains in its memory the best position it ever encountered. In the globalvariant of PSO the best position ever attained by all individuals of the swarm is communicated to all the particles. In the localvariant, each particle is assigned to a neighborhood consisting of a pre specified number of particles. In this case, the best position ever attained by the particles that comprise the neighborhood is communicated among them. Finally, the PSO algorithm maintains the best fitness value achieved among all particles in the swarm, called the global best fitness, and the candidate solution that achieved this fitness, called the global best position.

PSO also keeps the track of the all the best values that the particles have achieved so far. Each particle maintains its position, composed of the candidate solution and its evaluated fitness, and its velocity. It remembers the best fitness value it has achieved, referred to as the individual best position or individual best candidate solution.

The PSO algorithm consists of three steps, which are repeated until some stopping condition is met

1. Evaluate the fitness of each particle
2. Update individual and global best fitness and positions
3. Update velocity and position of each particle

A.Updation of velocity and particle position

The new particle positions are calculated by adding their previous position to their corresponding updated

velocity. Updating velocity is very important in PSO. The acceleration constants helps in improving the particle position by comparing with the previous particles position and also makes the particle to follow the best neighbour's direction :

The velocity of each particle in the swarm is

vi(t+1) = w vi (t) + c1 r1 [xˆ i (t) – xi (t)] + c2 r2 [g (t) – xi (t)]

and the position

$X_i^{k+1} = x_i^k + v_i^{k+1}$

where

$V_{i(t+1)}$ is the velocity of the particle at t+1th iteration

Vi(t) velocity of the particle at t iteration

$X_i(k+1)$ is the position of the particle

c1 and c2 are the acceleration factor constants related to gbest and pbest

r1 and r2 are the random numbers between 0 and 1

i=1,2,3…20

B.PSO parameters

Swarm size N=30

 Acceleration constants c1=1.5

Acceleration constant c2=1.5

Objective function- ITAE (Integral Time Absolute error)

f(x)=∫ t|e(t)|dt(8)

C.Flowchart

The flowchart given below represents the procedure of PSO algorithm implementation.
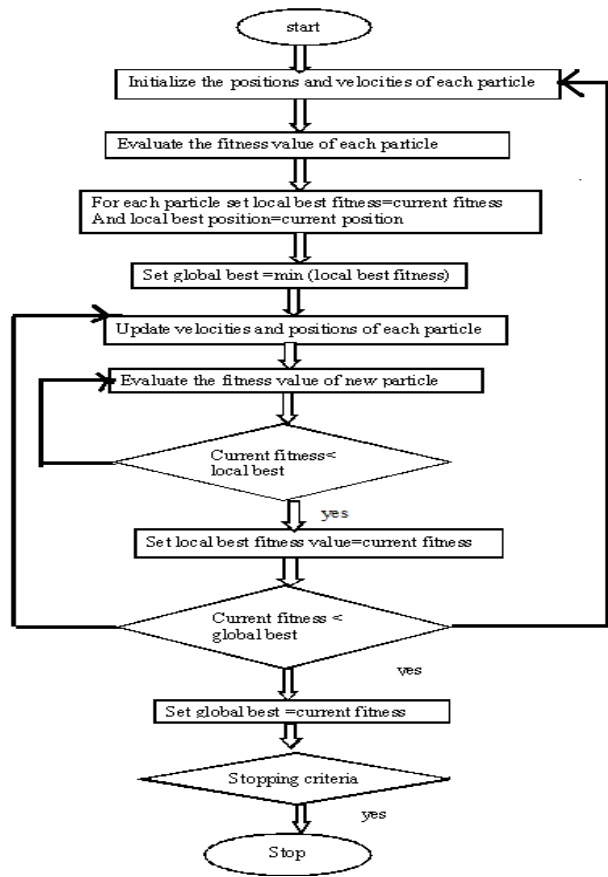
Figure 6: PSO flow chart

The performance of various control techniques of load frequency control of single area system is analysed in this work through simulation in the MATLAB Simulink environment. The Simulink diagram of ALFC single area test system with 2DOF-PID controller used in this work is shown in Fig.7
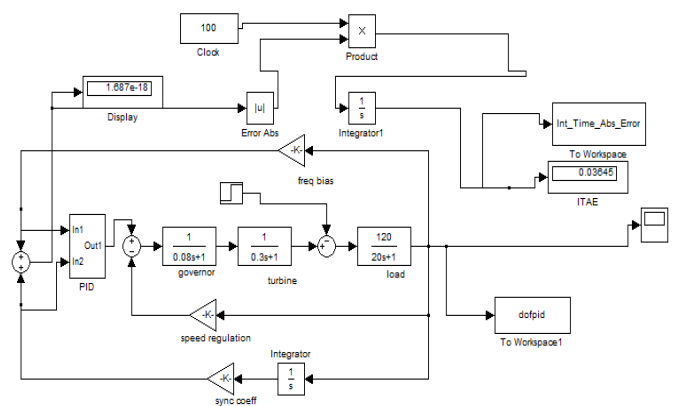
Figure 7. Simulink diagram of ALFC single area system

IV. IV RESULTS AND DISCUSSIONS

A. Case1: analysis of various controllers

Performance of PI, PID and 2DOF-PID controllers are compared with their settling time, overshoot,

undershoot and ITAE. The results are given in table I.

TABLE I. Comparison of various controllers

controller	Settling time(s)	Overshoot	Under shoot	ITAE
PI	0.125	0.205	-0.205	0.9599
PID	0.072	0.085	-0.145	0.1705
2DOF-PID	0.061	0.015	-0.02	0.03785
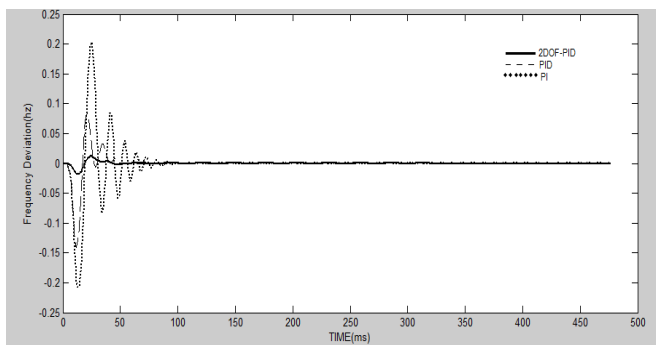
Figure 8. Comparison of various controllers

From the Fig.8 the proposed controller 2DOF-PID shows less ITAE, less settling time, less overshoot and less undershoot when compared to conventional controllers.

B. case 2: Increase of 10% and 20% Step load disturbance

Increase in Load disturbance of 10% and 20% (1.1pu and 1.2p.u) is given to the ALFC single area test system and its performance is evaluated and results are presented in table II and table III respectively.

controller	Settling time(s)	Overshoot	Under shoot	ITAE
PI	0.13	0.21	-0.215	0.9846
PID	0.079	0.085	-0.145	0.1778
2DOF-PID	0.063	0.018	-0.02	0.03924

TABLE II.10% step load disturbance
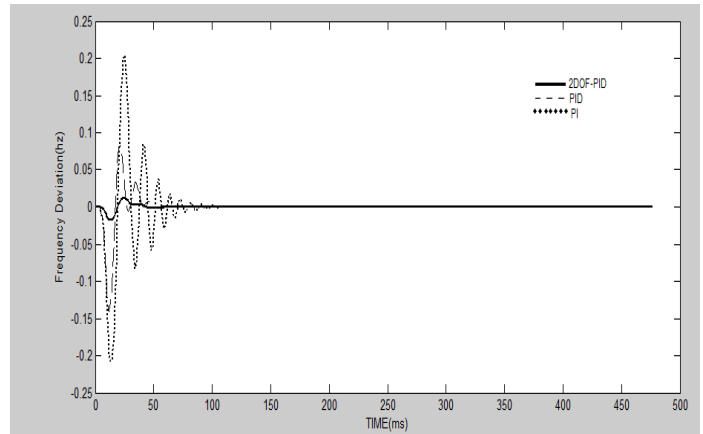
Figure 9. 10% step load disturbance

TABLE III.20% step load disturbance

controller	Settling time(s)	Overshoot	Under shoot	ITAE
PI	0.12	0.2	-0.2	0.9351
PID	0.073	0.082	-0.14	0.1632
2DOF-PID	0.064	0.015	-0.015	0.03645
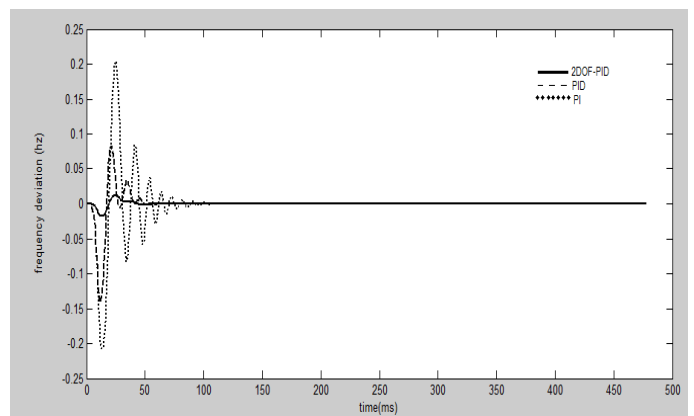
Figure 10. 20% step load Disturbance

Above Fig.9. & Fig.10, show that the 2DOF controller is able to reject disturbance for the increase of 10% and 20% step change in load demand with less ITAE and it exhibits better dynamic performances.

C. case 3: Analysis of 50% Parameter Uncertainty

The change in parameter of the system is considered to check the robustness of the proposed controller. The 2DOF-PID is robust for 50% parameter uncertainty in governor time constant (T_g). The result is presented in table IV

TABLE IV.Robustness analysis

controller	Settling time(s)	Overshoot	Under shoot	ITAE
PI	0.15	0.23	-0.22	2.016
PID	0.084	0.1	-0.15	0.2336
2DOF-PID	0.065	0.015	-0.0195	0.03711
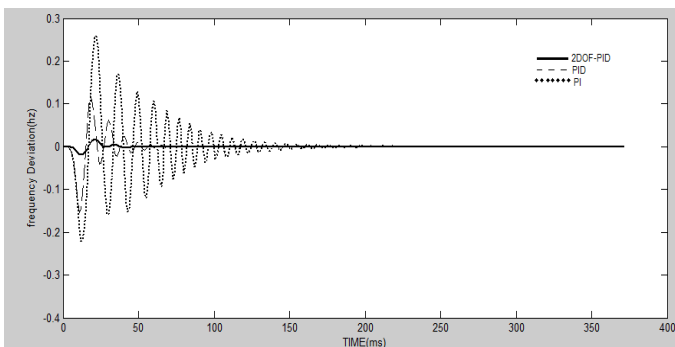
Figure 11.Robustness analysis

Thus the performance of 2DOF-PID seems to be better than conventional controller in terms of dynamic performance like settling time, overshoot, undershoot and performance index (ITAE).

IV.CONCLUSION

By comparing the simulation results of PI, PID, and 2DOF-PID controllers, the performances of the proposed controller proves to be better as it reduces the transient deviations and damp out the frequency for increase in 10% and 20% step load disturbance. Further the proposed controller is robust for ±50% change in governor time constant. It can be extended by changing turbine and generator time constant also. Thus, it is concluded that 2DOF-PID controller in all cases have lesser settling time, reduced performance index and with less overshoot and undershoot when compared to PI,PID controllers.

V. Acknowledgment

VI. References

[1] D.K.Chaturvedi, P.S.Satsangi , P.K. Kalra "Load frequency control:a generalized neural network approach" Electrical Power and Energy Systems 21(1999)405-415 Elsevier

[2] Dola Gobinda Padhan,Somanath Majhi " A new control scheme for PID load frequency controller for single area and multi area systems" ISA Transactions 52(2013)242-251 Elsevier

[3] O.I.Elgerd "Electric Energy System Theory;An introduction "Mc Graw Hill.1971

[4] M.K.Sherbiny El "Efficient fuzzy logic load frequency controller"Energy Conversion and Management 43(2002)1853-1863 Elsevier

[5] S.P.Ghosal "Optimizations of PID gains by particle swarm optimizations in fuzzy based automatic generation control"Electric Power Systems research 72(2004)203-212 Elsevier

[6] Muwaffaq Irsheid Alomoush "Load frequency cntrol and automatic generation control using fractional order controllers" Electr Engg(2010)91:357-3688 Springer

[7] Wen tan"tuning of PID load frequency controller for power systems" Energy Conversion and Management 50(2009(1465-1472 Elsevier

[8] V.soni.G.Parmar,M.Kumar and S.Panda "hybrid grey wold optimization-pattern search optimized 2DOF-PID controllers for load frequency control in interconnected thermal power plants"ICTACT journal on soft computing April 2016,volume 06, issue 03.

[9] RV Rao, VJ savsani,DP vakharia "Teaching learning based optimization: an optimization method for continuous non-linear large scale problems"Infinite science 2012;183(10:1-15

[10] P. Kundur."Power system stability and control",Tata Mc Graw Hill,2009.

[11] J.Talaq,A.I.Fadel and Basri,"Adaptive Fuzzy Gain Scheduling for Load Frequency control",IEEE Transactions on Power System,Vol.14.No.1,pp.145-150,1999

[12] D.P.KothariandI. J. Nagrath." Power System Engineering", 2nd edition,Tata

Design of CMOS 8-Bit Parallel Adder Energy Efficient Structure using SR-CPL Logic Style

Aravinth.T.S[1], Sundar.R[2], Felix Muthu[3]

[1]*Assistant Professor,* [2]*Assistant Professor,* [3]*PG Scholar*
[1,3] *Department of Electronics & Communication Engineering,*
[2]*Department of Electrical & Electronics Engineering*
[1,3]*Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, Tamil Nadu, India*
[2]*Karpagam College of Engineering, Coimbatore, Tamil Nadu, India*

ABSTRACT

We present high speed and low power 8-Bit parallel adder cells designed with modified SR-CPL logic styles that had a reduced power delay product (PDP) as compared to the previous logics DPL and pass transistor logic. We have carried out a comparison against other parallel adders reported as having a low PDP, in terms of speed and power consumption. All the parallel adders were designed with a 0.18µm CMOS technology virtuoso cadence environment. Simulations of the circuit show that the proposed parallel adders have reduced the power from 0.33mW to 0.24mW.

Keywords : SR-CPL logic Styles, PDP, DPL, Pass Transistor, Logic, Virtuoso Cadence Environment.

I. INTRODUCTION

Energy effectiveness is one of the most required features for modern low power electronic devices , which have been designed for high-performance portable applications. In one hand, the ever increasing market segment of portable electronic devices demands the availability of low-power building blocks that enable the implementation of long-lasting battery-operated systems. On the other hand, the general trend of increasing operating frequencies and circuit complexity, in order to cope with the throughput needed in modern high-performance processing applications, requires the design of very high-speed circuits. The power-delay product (PDP) metric relates the amount of energy spent during . Thus, taking this fact into consideration, the design of a Parallel adder having low-power consumption and low propagation delay results of great interest for the implementation of modern digital systems. In this paper, we report the design and performance comparison of Parallel-adder cells implemented with an alternative internal logic structure, based on the multiplexing of the Boolean functions XOR/XNOR and AND/OR, to obtain balanced delays in SUM and CARRY outputs, respectively, and SR-CPL logic styles, in order to

the realization of a determined task, and stands as the more fair performance metric when comparing optimizations of a module designed and tested using different technologies, operating frequencies, and scenarios. Addition is a fundamental arithmetic operation that is broadly used in many VLSI systems, such as applicationspecific digital signal processing (DSP) architectures and microprocessors. This module is the core of many arithmetic operations such as addition/subtraction, multiplication, division and address generation. As stated above, the PDP exhibited by the Parallel-adder would affect the system's overall performance

reduce power consumption. The resultant Parallel-adders show to be more efficient on regards of power consumption and delay when compared with other ones reported previously as good candidates to build low-power arithmetic modules. This paper is organized as follows. Section II presents the internal logic structure adopted as standard in previous papers for designing a Parallel-adder cell. Section III introduces the alternative internal logic structure and the SR-CPL logic styles used to build the proposed Parallel-adders. Section IV

explains the features of the simulation environment used for the comparison carried out to obtain the power and speed performance of the Parallel-adders. Section V reviews the results obtained from the simulations, and Section VI concludes this work.

II. Literature Review

Recursive Approach to the Design of Parallel Self-Timed Adder

It proposes to use a zero-delay overhead self-timed pipeline style that supports very high speed operation. Developed techniques to enable the application of zero delay-overhead self-timed pipeline in this context and realize run-time pipeline depth control. Simulations under variable data rate scenarios demonstrate a significant performance gain [1].

CMOS Full Adder for Energy Efficient Arithmetic Applications

The design of high-speed low-power full adder cells based upon an alternative logic approach has been presented. Which results in a great improvement on regards of power-delay metric for the proposed adders, when compared with several previously published realizations [2].

Area Efficient Self Timed Adders for Low Power Applications in VLSI - The parallel asynchronous self-timed adder circuit is efficiently described using a handshaking protocol and also compared with other adders proposed adders. The MAC unit is implemented and the process is achieved efficiently. Simulation results demonstrate the effectiveness of the proposed framework in parallel prefix adder using multiplication through addition process [3].

Design of Low Power Asynchronous Parallel Adder - This paper describes an asynchronous parallel adder. It is based on Radix method for faster computation of sum and to reduce delay caused by carry chain. The computation has been carried out using parallel process. The aim of this work is to reduce the Power Delay Product (PDP) and Energy Delay Product (EDP) of an adder [3].

Self-Repetitive Approach to the Design of a Parallel

Self- Timed Adder - This brief presents a parallel single-rail self-timed adder. It is based on a recursive formulation for performing multi bit binary addition. The operation is parallel for those bits that do not need any carry chain propagation [4].

Design of a Parallel Self-Timed Adder Using Recursive Approach - This brief presents a parallel single-rail self-timed adder. It is based on a recursive formulation for performing multibit binary addition. The operation is parallel for those bits that do not need any carry chain propagation. Thus, the design attains logarithmic performance over random operand conditions without any special speedup circuitry or look-ahead schema [5].

III. The Existing Method
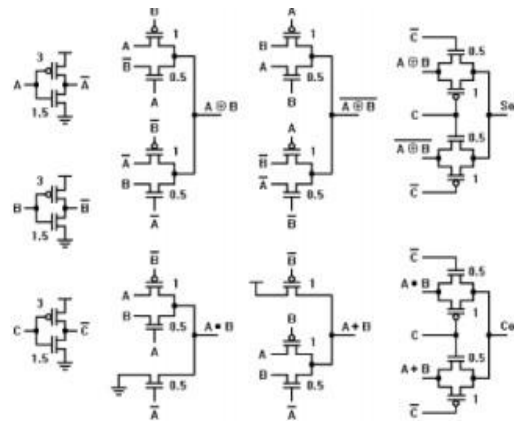
Figure 1 Existing Circuitary
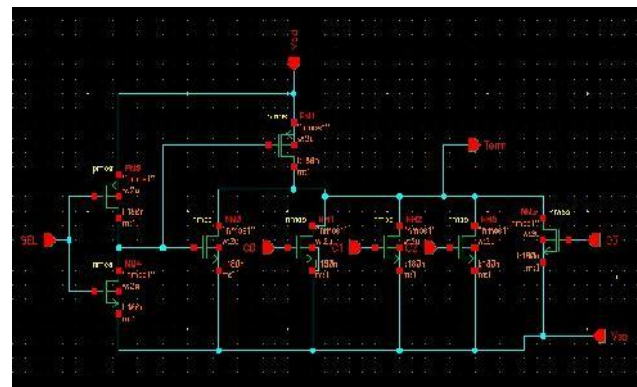
Figure 2 Completion Detection Circuit

Many projects have been published regarding the optimization of low-power full-adders, trying different options for the logic style standard CMOS , differential cascaded voltage switch (DCVS) , complementary pass-transistor logic (CPL) , double pass transistor logic (DPL) , swing restored CPL (SR-CPL) and hybrid styles , and the logic structure used to build the adder

module. The internal logic structure has been adopted as the standard configuration in most of the enhancements developed for the 1-bit full-adder module. In this configuration, the adder module is formed by three main logical blocks: a XOR-XNOR gate and XOR blocks or multiplexers to obtain the SUM (So) and CARRY (Co) outputs. The major problem regarding the propagation delay for full-adder built with the logic structure is that it is necessary to obtain an intermediate .8_signal and its complement, which are then used to drive other blocks to generate the final outputs. Thus, the overall propagation delay and, in most of the cases, the power consumption of the full-adder depend on the delay and voltage swing of the .8_signal and its complement generated within the cell. So, to increase the operational speed of the full-adder, it is necessary to develop a new logic structure that does not require the generation of intermediate signals to control the selection or transmission of other signals located on the critical path.

IV. The Proposed Method

In this method report there are not signals generated internally that control the selection of the output multiplexers. Instead, the input signal, exhibiting a full voltage swing and no extra delay, is used to drive the multiplexers, reducing so the overall propagation delays. The capacitive load for the input has been reduced, as it is connected only to some transistor gates and no longer to some drain or source terminals, where the diffusion capacitance is becoming very large for sub-micrometer technologies. Thus, the overall delay for larger modules where the signal falls on the critical path can be reduced. The propagation delay for the So and Co outputs can be tuned up individually by adjusting the XOR/XNOR and the AND/OR gates; this feature is advantageous for applications where the skew between arriving signals is critical for a proper operation and for having well balanced propagation delays at the outputs to reduce the chance of glitches in cascaded applications. The inclusion of buffers at the full-adder outputs can be implemented by interchanging the XOR/XNOR signals, and the AND/OR gates to NAND/NOR gates at the input of the multiplexers, improving in this way the performance for load-sensitive applications.
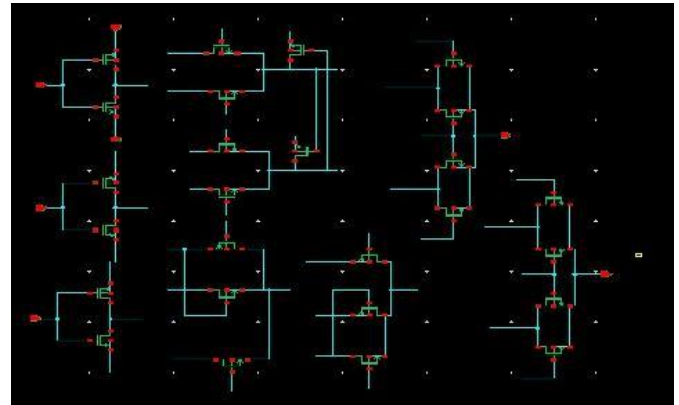
Figure 3 SR-CPL Logic Style

V. Energy Efficient Parallel Full Adder

After studying the truth table of full adder in Table 1, it can be observed that the So output is equal to the (A XOR B) value when C=0, and equal to (A XNOR B) when C=1. From this observation we conclude that a multiplexer will be used to obtain the respective value based upon the Carry input, as stated earlier. Using the same scenario, the Co output is equal to the (A AND B) value when C=0, and (A OR B) value when C=1. In the similar way, carry will be used to drive a multiplexer. Hence, an energy efficient logic scheme to design a full adder cell can be formed by a logic block to get the (AXOR B) and (AXOR B) signals, other block to obtain (A AND B) and (A OR B) signals, and two multiplexers being driven by the Carry input to generate the So and Co outputs, as shown in Figure 2.
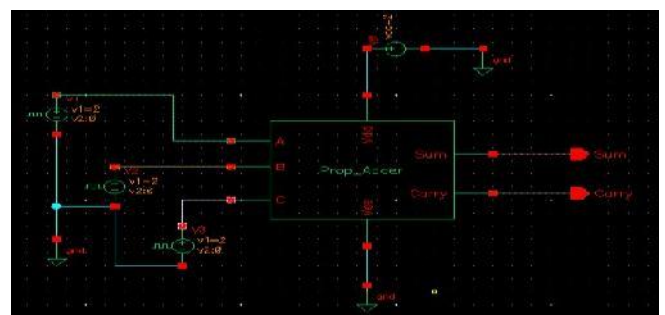
Figure 4. Proposed Parallel Half Adder using SR-CPL

VI. Results
Table 1 Current Dissipation

Transitions	Design 1 (D1)	Design 2 (D2)	Reference (CPL)
A 0→1→0, B=0, Cin =1	32.6u	32.5u	47.2u
A 0→1→0, B=1, Cin =0	32.1u	31.4u	47.4u
B=0→1→0, A =0, Cin =1	32.5u	31.3u	45.2u
B 0→1→0, A=1, Cin=0	35.5u	32.7u	45.4u
Cin 0→1→0, A=0, B=1	27.7u	24.7u	45.2u
Cin 0→1→0, A=1, B=0	31.0u	25.1u	45.2u

Table 2 Delay

Transitions	Design 1 (D1)	Design 2 (D2)	Reference (CPL)
A 0→1→0, B=0, Cin =1	361.2p	239.2p	330.2p
A 0→1→0, B=1, Cin =0	391.7p	241.2p	304.2p
B 0→1→0,A =0, Cin =1	386.6p	240.8p	288.8p
B 0→1→0, A=1, Cin=0	383.1p	229.4p	322.3p
Cin 0→1→0, A=0, B=1	329.8p	205.6p	299.4p
Cin 0→1→0, A=1, B=0	368.1p	188.5p	313.2p

Table 3 Power Delay Product

Transitions	Design 1 (D1)	Design 2 (D2)	Reference (CPL)
A 0→1→0, B=0, Cin =1	21.2	13.9	28.1
A 0→1→0, B=1, Cin =0	22.6	13.6	25.9
B=0→1→0,A =0, Cin =1	22.6	13.5	23.6
B 0→1→0, A=1, Cin=0	24.5	13.5	26.3
Cin 0→1→0, A=0, B=1	16.5	9.2	24.5
Cin 0→1→0, A=1, B=0	20.5	8.5	25.5

VII. Discussion
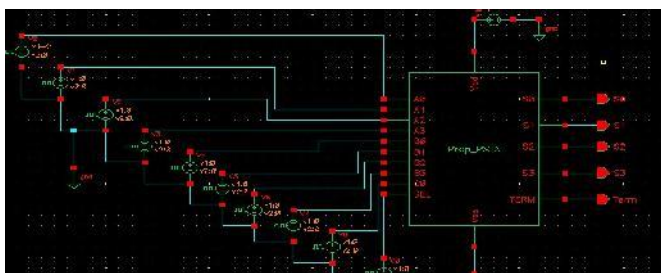
Figure 5. 8-Bit Parallel Full Adder using SR-CPL

Two new designs based on SR-CPL and DPL style full adders are being examined in this project. The main advantages of this design are: Multiplexers are directly controlled by Cin instead of internally generated signals thereby reducing delay.
Capacitive load on Cin is reduced.
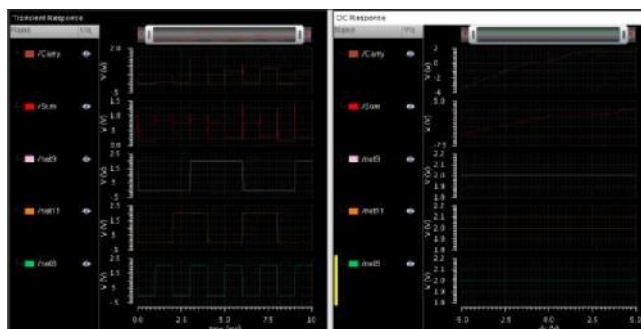
Figure 6 Output Plot of Proposed Circuitry

The propagation delay of So and Co can be tuned by sizing XOR/XNOR gates appropriately. The inclusion of buffer at input can be integrated by using NAND/NOR gates instead of XOR/XNOR gates. Buffers are placed at the inputs are placed to account for the load the

device offers at the inputs.

Also, since the designs presented here consist of pass transistor logic which has no direct power supply connection, the power consumed by the device also comes through these inverters. The output inverters account for the power due to degraded voltage swing and slopes of full adder output. The full adders have been simulated using 180-nm CMOS technology using cadence virtuoso. The value of supply voltage VDD used was 1.8V.
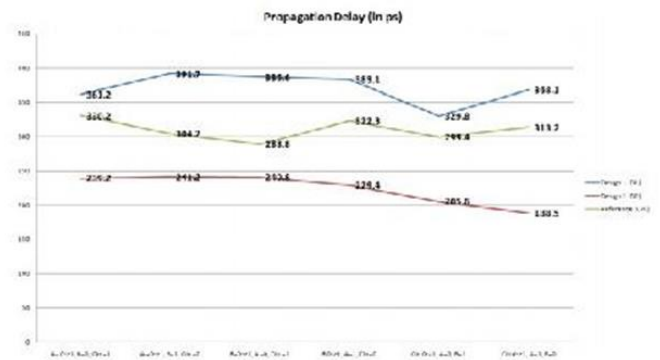
Figure 7 Power Delay Plot for SR-CPL vs Existing Styles
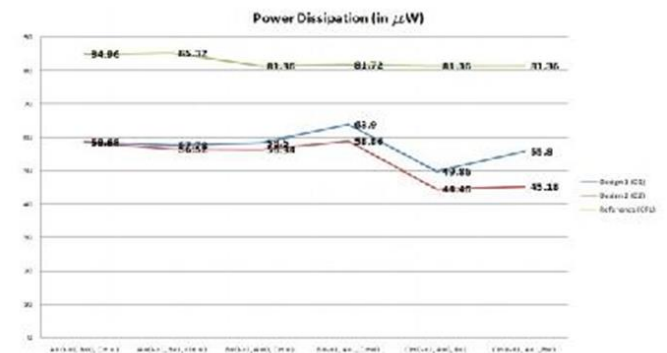
Figure 8 Power Dissipation Plot for SR-CPL vs Existing Styles
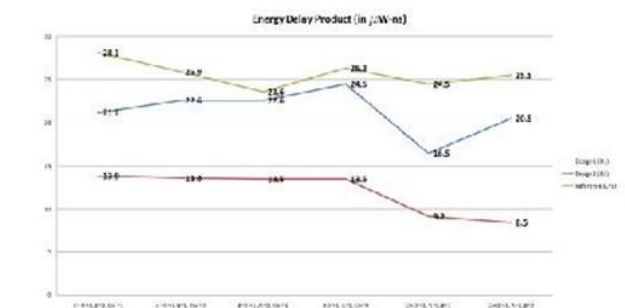
Figure 9 Energy Delay Product Plot for SR-CPL vs Existing Styles

VIII. Conclusion

We have presented SR-CPL parallel full adder and CPL style adder. The key features observed were: 1. the proposed designs reduce both total average power and worst case delay of the circuit. Total average power reduction of about 30-38% is observed for D1 and about 27-45% for D2 2. Delay of D1 is comparable to that of the CPL logic (it is slightly greater for most transitions). Delay of D2 is significantly smaller as compared to the reference CPL Design – from about 16% to 31% 3. The overall Power delay product of both the designs is reduced as compared to the standard CPL design ranging from about 5% to 25% for D1 and about 43% to 66% for D2 4. The designs are more efficient both power wise and delay wise as compared to the standard CPL design used 5. The transistor count for the proposed designs is also much less as compared to the s CPL adder (26 & 28 as compared to 38) 6. The proposed designs occupy much less area as compared to the CPL adder (116 μm2 & 118 μm2 for D1 and D2 vs 238 μm2 for CPL adder)

IX. References

[1] A. M. Shams and M. Bayoumi, "Performance evaluation of 1-bit CMOS adder cells ", IEEE ISCAS, Orlando, Florida, May 1999, pp. I27 -130.

[2] A.P.Chandrakasan, S.Sheng and R.W.Brodersen, "Lowpower CMOS digital design ", IEEE JSSC, Vol. 27, April 1992, pp. 473-483.

[3] N. Weste and K. Eshraghian, Principles of CMOS design, A system perspective, Addison-Wesley, 1988. 4K. M. Chu and D. Pulfrey, "A comparison of CMOS circuit techniques: Differential cascode voltage switch logic versus conventional logic, "IEEE J. Solid-State Circuits, vol. SC-22, no. 4, pp.528-532, Aug.1987.

[4] K. Yano, K. Yano, T. Yamanaka, T. Nishida, M. Saito, K. Shimohigashi, and A. Shimizu, "A 3.8 ns CMOS 16-b multiplier using complementary pass-transistor logic," IEEE J. Solid-State Circuits, vol. 25, no. 2, pp. 388-395, Apr. 1990.

[5] M. Suzuki, M. Suzuki, N. Ohkubo, T. Shinbo, T. Yamanaka, A. Shimizu, K. Sasaki, and Y. Nakagome, "A 1.5 ns 32-b CMOS ALU in double pass-transistor logic," IEEE J. Solid-State Circuits, vol. 28, no. 11, pp. 1145-1150, Nov.1993.

[6] R. Zimmerman and W. Fichtner, "Low-power logic styles: CMOS Versus pass-transistor logic," IEEE J. Solid-State Circuits, vol. 32, no. 7, pp. 1079-1090, Jul. 1997.

[7] N. Zhuang and H. Wu, "A new design ofthe CMOSfull adder", IEEE JSSC, Vol. 27, No. 5, May 1992, pp. 840-844

[8] A. M. Shams and M. Bayoumi, "A new cellfor low power adders ", Proceedings of the International MWSCAS, 1995.

[9] C. Chang, J. Gu, and M. Zhang, "A reviewof 0.18- full adder performances for tree structured arithmetic circuits," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 13, no. 6, pp. 686-695, Jun. 2005.

[10] M. Aguirre and M. Linares, "An alternative logic approach to implement high-speed low-power full adder cells," in Proc. SBCCI, Florianopolis, Brazil, Sep. 2005, pp.166-171.

[11] Reto Zimmermann and Wolfgang Fichtner, Fellow, IEEE" Low-Power Logic Styles: CMOS Versus Pass- Transistor Logic" in IEEE Journal Of Solid-State Circuits, Vol. 32, No. 7, July 1997.

[12] Aguirre-Hernandez, Mariano, and Monico Linares- Aranda. "CMOS fulladders for energy-efficient arithmetic applications." Very Large Scale Integration (VLSI) Systems, IEEE Transactions on 19.4 (2011): 718-721.

[13] Quintana, J. M., et al. "Low-power logic styles for full- adder circuits."Electronics, Circuits and Systems, 2001. ICECS 2001. The 8th IEEE International Conference on. Vol.3. IEEE, 2001

[14] J Rabaey, A Chandrakasan, and B Nikolic, Digital Integrated Circuits: A Design Perspective. Upper Saddle River, NJ: Prentice-Hall, 2003 4. Vijay, V., J. Prathiba, and S. Niranjan Reddy. "A REVIEW OF THE 0.09 μ m STANDARD FULL ADDERS." International Journal of VLSI Design & Communication Systems 3.3 (2012)

Optimal Polling Point Selection and Channel Allotment Scheme for Clustered WSN

T. Parimalam * [1], K. Nirmala[2]

*[1]HOD in Computer Science, Nandha Arts and Science College,Erode,Tamil Nadu,India
pari.phd12@gmail.com[1]
[2]Research scholar,Nandha Arts and Science College,Erode, Tamil nadu,India
nirmalabuasc@gmail.com[2]

ABSTRACT

Wireless Sensor Networks are constructed to manage the environment surveillance operations. The mobile collector performs the data collection operations at the polling points. The polling point selection is carried out with the cluster head pair details. The data forwarding operations are performed with the distributed load balanced clustering and dual data uploading (LBC-DDU) method. Optimal polling point selection scheme is constructed in the Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) method. The bandwidth assignment tasks are managed with the channel allotment scheme. Network area verification and rescheduling process manage the spatial coverage optimization for the data collection process. The data gathering scheme increases the network lifetime with minimum query response delay. Network traffic levels are minimized in the data collection framework.

Keywords: Wireless sensor networks (WSNs), data collection, multi-user multiple-input and multiple –output (MU-MIMO), mobile control, virtual MIMO.

I. INTRODUCTION

Wireless sensor networks (WSNs) are application specific networks composed of small nodes, which can sense the environment, collect the data, do aggregation and every single node can communicate with each other wirelessly via radio link [1]. Today's fast technology improvements in low-power and wireless communication have provided a good condition for WSNs in real-world applications and distributed sensor applications have increased significantly such as wild life and ocean life monitoring, supervising the vibration of structures, automatic warning, supervising the agricultural applications and target tracking.

Nodes have limitations in memory, process and energy; therefore it is difficult to design WSNs. Among the above-

mentioned limitations, energy is the most important one because when the sensors are installed their batteries cannot be replaced or charged. Thus, energy considerations are the most prominent factors in WSNs routing. One of the most famous routing algorithms for WSNs is clustering-based hierarchical routing. In this method, all nodes are divided into groups called clusters based on specific methods. In each cluster, one node is selected as a cluster head (CH) and other nodes are considered as normal nodes. Different parameters are taken into account while selecting a CH in various methods. In the major part of clustering algorithms, the main goal is to achieve uniform energy distribution to increase network lifetime. In this type of routing, sensor nodes play different roles and they may have different energy consumption according to their role.

This group of methods is the best class of routing algorithms for WSNs. A CH is able to manage and schedule intra-cluster activities, and as a result nodes may change their state to low-power sleep mode and

reduce energy consumption. The nodes might be utilized in a round-robin order and a specific time could be determined for data transmission and receive. Retransmitting is avoided and data redundancy in target region is decreased and medium access collision is also avoided. Nodes alternatively transmit their data to CH. The CH collects the data, compresses it and transmits the compressed data to the base station (BS). Transmission to the BS might be done directly or in a multi-hop process and with the incorporation of other CHs. Since CHs transmit their data over longer distance, the energy consumption rate is higher in CHs. A simple approach to balance energy consumption is to reselect the CHs periodically. In this case, the role of CH is changed. The structure between normal nodes, CHs and BS might be repeated.

II. RELATED WORK

With the emergence of mobile sensors, extensive researches have been promoted on target coverage of WSNs. According to different application scenarios, the existing studies can be classified into three categories: (1) route patrol for collecting data from fixed targets [10], (2) detection of mobile targets [4], [5], [3] and (3) target coverage in dynamic environments [6]. In these studies, mobile sensors move actively to improve the surveillance quality, but the optimization of sensor movement is not explicitly considered. Reactive mobility is exploited to improve the quality of target detection, but the movement of sensors is not considered as the primary optimization objective. Mobile sensors are scheduled to replace failed static sensors in order to guarantee coverage ratio with minimum movement distance. But each sensor concerned cover only one target and the maximum moving distance for each mobile sensor is limited. An optimal velocity schedule is proposed to minimize energy consumption in movement when the road condition is uniform.

Many research efforts have also been made to improve the area coverage with mobile sensors with the aim of maximizing the covered area. In [7], Voronoi diagrams are used to detect coverage holes. Sensors are dispatched to cover the detected holes. As a result, the area coverage ratio is improved. Further, a multiplicatives weighted Voronoi diagram is used to discover the coverage holes corresponding to different sensors with different sensing ranges [8].Voronoi diagram to discover

the coverage holes corresponding to different sensors with different sensing ranges. Voronoi diagrams in these studies are constructed according to the position of mobile sensors, and thus need to be recomputed after each round of sensor movement. In [9], mobile sensors are used to improve energy efficiency of sensors in area coverage. When destinations have been determined, mobile sensors are designed to move along the shortest path to minimize the energy consumption.

Given designated destinations, k-coverage is studied. In this work, a competition scheme is proposed to minimize energy consumption in movement. Recently, parameterized algorithms were exploited to find maxlifetime target coverage [11] and min-power multicast paths WSNs. In these studies, destinations of mobile sensors are given in advance and the energy efficiency is considered in the path finding process.

Mobility of sensors could also be exploited to enhance network connectivity after the coverage stage is completed. A triangular deployment strategy is proposed to dispatch sensors to connect the network after deploying mobile routers to maximize the coverage area. In the proposed strategy, sensors move along the shortest path to the corresponding triangular vertices in order to save energy [2]. The authors considered a hybrid network consisting of both static and mobile sensors. It first divides the static sensors into groups as large as possible and then seeks the minimum number of mobile sensors to connect these static sensor groups. A sensor node relocation approach is proposed to maintain connectivity between a region of interest and a center of interest outside the deployment region where a particular event happens.

The originality of this study and differences from the existing work include. (1) In this work, sensors move reactively and each sensor can cover more than one target, which is more general in practice, but also makes the problem more complicated. (2) The Voronoi diagram of targets is adopted to find the nearest sensor, which avoids blind competition among mobile sensors. Besides, because our solution generates the Voronoi diagram according to the position of targets, it does not require re-computation of the Voronoi diagram as the targets are static. This contributes to the lower complexity of the proposed solution. (3) Destinations of mobile sensors are unknown, which should be computed by our algorithms. When mobile sensors move to these destinations, both

target coverage and network connectivity are satisfied. (4) In order to investigate the impact of network parameters on the performance of our algorithms, analyses and evaluations are given according to the simulation experiment results, which provides a reference for practical engineering and theoretical basis for the design of mobile sensor networks.

III. SYSTEM METHODOLOGY

3.1 Data Gathering using Joint Virtual MIMO

The joint virtual MIMO and data gathering (vMDG) model is adapted to schedule the nodes. There are n wireless nodes randomly distributed in a planar field. Each node u knows its position denoted by (X_u, Y_u). Each node is equipped with an antenna and can adjust its transmission power arbitrarily. Four important assumptions are considered for problem definition. First, it is assumed that the network is synchronized. Despite some concerns over the infeasibility of the MIMO mode due to lack of simultaneous synchronization, some works a small synchronization error didn't greatly decrease the performance of vMIMO transmission. For example, Nguyen et al. analyzed that the performance degradation increased with the synchronization error and the number of cooperative transmission and reception nodes [12]. The cooperative MIMO system was rather tolerant for small ranges of synchronization error and the degradation was negligible for synchronization error range as small as $0.2T_s$, where T_s was the symbol period. As a result, it is reasonable to assume perfect time synchronization among the network. Second, transmission collision can be avoided. In general, MAC protocols based on adaptive modulation can be used and in particular, the MAC protocol uses a variable-length energy-minimizing TDMA scheme for interference avoidance also used.

Third, this work ignores the cost of sharing the control information for vMIMO transmission in the data gathering. This assumption is based on the following reasons. The control packet is relatively short compared with the data packet. Second, the algorithm will construct a tree. Most of the links in this tree are not long. The energy consumption of additional data exchange procedure will not greatly impact the energy consumption of vMIMO communication. Fourth vMIMO involving two (2) nodes at most based on the

following rationale. Sajid et al. have given two conclusions by simulations. The SIMO, MISO and MIMO models are more energy-efficient compared with the SISO mode. Second, the 2 * 2 mode is much more energy-efficient than 1 * 1 (SISO), 3 * 3 (MIMO) and 4 * 4 (MIMO) modes. If more nodes cooperate with others in vMIMO transmission, the management of the cooperative nodes will also become more complex. The vMIMO transmission includes four different communication modes, SISO (1 * 1), SIMO (1 * 2), MISO (2 * 1) and MIMO (2 * 2) respectively. In the SIMO, MISO and MIMO modes, the transmitter or receiver sides may contain two nodes. For convenience, two nodes in one transmitter or receiver side will be referred to as a cooperative node pair.

The joint vMIMO and data gathering (vMDG) problem is to select a set of cooperative node pairs, construct a vMIMO-aware topology and perform vMIMO-aware routing on the topology, so that all nodes will send their sensor data to the base station with vMIMO transmissions. The optimization objective of this problem is to minimize the total energy consumption of data gathering for wireless sensor networks.

IV. SYSTEM IMPLEMENTATION

4.1. Optimal Polling Point Selection Scheme

The polling points are identified to locate the mobile collectors. The mobile collector performs the data gathering process on the selected polling points. The mobile collector makes the communication with the cluster head pairs from the polling points. The clusters are formed with reference to the node location and resource information. The cluster heads are elected with reference to the resource and coverage level information. Cluster head pairing process is initiated to group up the nearest clusters for data collection process. The polling point selection process is integrated with the cluster head pair selection process. Partitioning method is applied to fetch optimal polling points. The vMIMO scheduling scheme is tuned to handle multiple cluster based data collection. The bandwidth allocation is carried out for the mobile collection data communication process.

4.2. Spatial Coverage Analysis

The sensor coverage information is categorized into two types. They are sensing coverage and transmission coverage. The sensing coverage deals with the data monitoring area levels. The transmission coverage indicates the data transmission distance levels. The spatial coverage analysis verifies both the sensing coverage and transmission coverage values. Sensor node deployment is carried out with the coverage values.

The network coverage reflects the entire monitoring area measurements. Data values are sensed and stored in the local storages. The data transmission is carried out through the cluster head to the mobile collector. In the sensing coverage verification the network is compared with the sensing coverage levels. The transmission coverage analysis includes the data transmission process between the node to cluster head and cluster head pair to the mobile collector. Node membership under the clusters is verified for the node to cluster head communication. The mobile collector trajectory plan is composed with reference to the cluster head pairing and polling point selection information. Cluster head pair communication is verified with the mobile collector receive ranges. The cluster assignment and head selection operations rescheduled with the spatial coverage information.

V. PERFORMANCE ANALYSIS

The mobile collector based data distribution scheme for Wireless Sensor Networks (WSN) is designed to perform data collection process. The sensor nodes are grouped with reference to the resource and proximity values. The cluster heads are selected to manage the data transfer process. A three tier framework is constructed to manage data transmission process. The mobile collector collects the data from the cluster heads. The cluster head collects the data values from the sensor nodes. Cluster head pair selection and pooling point selection tasks are carried out for the data collection process. The optimal polling point selection scheme is adapted in the polling point selection process.

The Load Balanced Clustering with Dual Data Upload (LBC-DDU) scheme is employed to perform the cluster head and pooling point selection process. The Multi User Multi Input and Multi Output scheduling scheme is

adapted for the bandwidth scheduling process. The Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) scheme is designed to handle the spatial coverage optimization, optimal polling point selection and channel allotment process. The system is tested with three performance measures. They are average response time, traffic rate and network lifetime parameters. The average response time measures the time period taken for the data collection process. Figure 5.1.shows the average response time analysis between the Load Balanced Clustering with Dual Data Upload (LBC-DDU) and Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) schemes. The analysis result shows that the Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) scheme reduces the average response time 30% than the Load Balanced Clustering with Dual Data Upload (LBC-DDU) scheme.
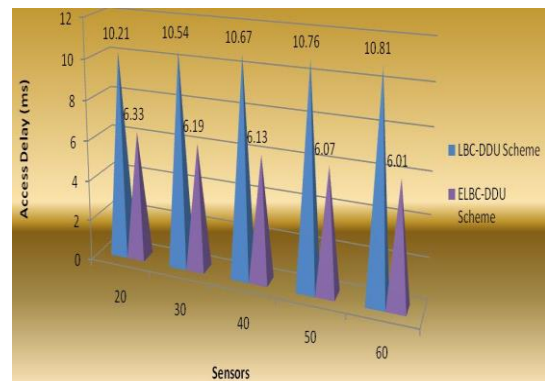
Figure 1. Average Response Time Analysis between (LBC-DDU) and (ELBC-DDU) schemes.

The traffic rate analysis is carried out to measure the message transmission level for the data transmission process. Figure 5.2.shows the traffic rate analysis between the Load Balanced Clustering with Dual Data Upload (LBC-DDU) and Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) schemes. The analysis result shows that the Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) scheme reduces the traffic rate 20% than the Load Balanced Clustering with Dual Data Upload (LBC-DDU) Scheme.
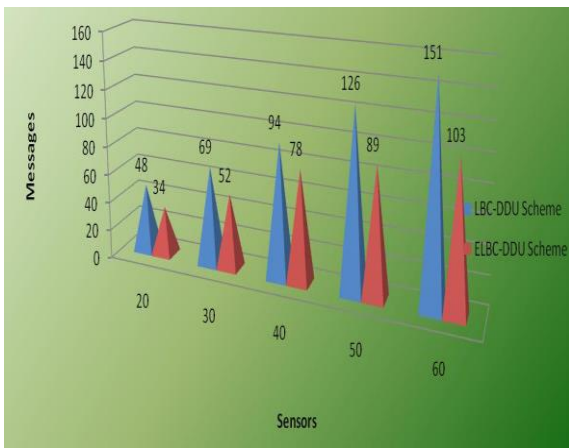
Figure 2. Traffic Rate Analysis between (LBC-DDU) and (ELBC-DDU) schemes
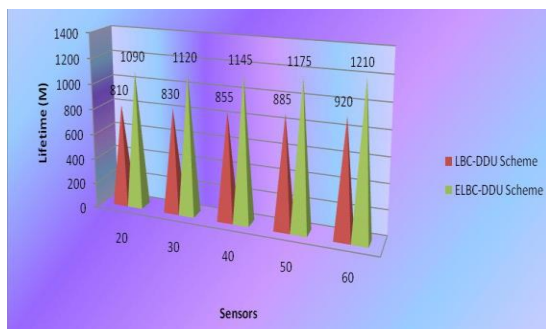
Figure 3. Network Lifetime Analysis between (LBC-DDU) and (ELBC-DDU) schemes.

The network lifetime analysis is carried out to measure the network lifetime for the WSN. Figure 5.3.shows the network lifetime analysis between the Load Balanced Clustering with Dual Data Upload (LBC-DDU) and Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) schemes. The analysis result shows that the Enhanced Distributed Load Balanced Clustering with Dual Data Upload (ELBC-DDU) scheme increases the network lifetime 35% than the Load Balanced Clustering with Dual Data Upload (LBC-DDU) scheme.

VI. CONCLUSION AND FUTURE ENHANCEMENT

The environment monitoring operations are managed with the sensor networks. The data capture and transmission operations are managed with mobile collectors. The mobile collector data transmission process is performed with the Distributed Load Balanced Clustering and Dual Data Uploading (LBC-

DDU) scheme. The Enhanced Distributed Load Balanced Clustering and Dual Data Uploading (ELBC-DDU) method is build with optimal polling point selection and channel allotment scheme. The ELBC-DDU also supports spatial coverage optimization to cover the entire network area. The network lifetime is increased with minimum energy consumption levels. Response time and traffic levels are minimized in the data collection process. The system can be enhanced with the following features. The system can be improved to handle malicious and anonymous request attacks. The data collection scheme can be tuned to handle mobile sensor environment.

VII. References

[1] Saman Siavoshi, Mehdi Tarhani and Habib Falari, "Geographical multi-layered energy-efficient clustering scheme for ad hoc distributed wireless sensor networks", IET Wireless Sensor Systems, Vol. 6, 2016.

[2] G. Blumrosen, B. Hod, T. Anker, D. Dolev and B. Rubinsky, "Enhancing rssi-based tracking accuracy in wireless sensor networks," ACM Trans. Sens. Netw., vol. 9, no. 3, pp. 1-29, 2013.

[3] X.-Y. Liu, K.-L. Wu, L. Kong and M.-Y. Wu, "Mobility increases the surface coverage of distributed sensor networks," Comput. Netw., vol. 57, no. 11, 2013.

[4] B. Liu, O. Dousse, P. Nain and D. Towsley, "Dynamic coverage of mobile sensor networks," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 2, Feb. 2013.

[5] G. Wang, M. Z. A. Bhuiyan, J. Cao and J. Wu, "Detecting movements of a target using face tracking in wireless sensor networks," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 4, pp. 939-949, 2014.

[6] S. He, J. Chen, X. Li, X. Shen and Y. Sun, "Cost-effective barrier coverage by mobile sensor networks," in Proc. IEEE 31st Annu. Int. Conf. Comput. Commun., 2012.

[7] H. Mahboubi, K. Moezzi, A. G. Aghdam and K. Sayrafian-Pour, "Self deployment algorithms for field coverage in a network of nonidentical mobile sensors," in Proc. IEEE Int. Conf. Commun., 2011, pp. 1-6.

[8] H. Mahboubi, K. Moezzi, A. G. Aghdam and K. S. Pour, "Distributed deployment algorithms for efficient coverage in a network of mobile sensors with nonidentical sensing capabilities," IEEE Trans. Veh. Technol., 2014.

[9] Y. Yang, M. I. Fonoage and M. Cardei, "Improving network lifetime with mobile wireless sensor networks," Comput. Commun., vol. 33, no. 4, 2010.

[10] L. He, J. Pan and J. Xu, "A progressive approach to reducing data collection latency in wireless sensor networks with mobile elements," IEEE Trans. Mobile Comput., vol. 12, no. 7, pp. 1308-1320, Jul. 2013.

[11] W. Luo, J. Wang, J. Guo and J. Chen, "Parameterized complexity of max-lifetime target coverage in wireless sensor networks," Theoretical Computer Science, vol. 518, pp. 32-41, 2014.

[12] Hongli Xu, Liusheng Huang and Yu-e Sun, "Joint Virtual MIMO and Data Gathering for Wireless Sensor Networks", IEEE Transactions On Parallel And Distributed Systems, Vol. 26, No. 4, April 2015.

A Survey on Deduplication Workload Resource for Big Data Applications

[1]V. Manochitra, [2]B.Jackline Jose
[1]Department of Information Technology, Bon Secours College for Women, Thanjavur
manokavishna@gmail.com
[2]Department of Information Technology, Bon Secours College for Women, Thanjavur
jackholyangel24@gmail.com

ABSTRACT

Deduplication seems to be a appropriate explanation for data detonation in the big data era by 1) slowing down the data growth speed by removing redundant data, and 2) relieving pressure on disk bandwidth by removing dismissed IO accesses. However, deduplication also introduces above to the system. For example, hash indexing needs be performed for every IO request to classify duplicates, which results in slower IO response time. In addition, extra CPU control is required to compute the hash values in each IO request, which leads to progressive vigor consumption. Since the capacity of IO needs is enormous and increasing in big data workloads, the overall performance and energy capability below different deduplication configurations is valuable to be deliberate methodically.

Keywords: Big Data, Deduplication, Hash Indexing, Resource Allocation, Big Data Analysis

I. INTRODUCTION

The demand for data storage and processing is increasing at a rapid speed in the big data era. Such a tremendous amount of data pushes the limit on storage capacity and on the storage network. The data analysis of the International Data Corporation (IDC), the volume of data in the world will reach 40 trillion gigabytes in 2020. In order to reduce the burden of maintaining big data, more and more enterprises and organizations have chosen to outsource data storage to cloud based big data storage providers. This makes data management a critical challenge for the big data storage providers. To achieve optimal usage of storage resources, many existing cloud storage providers perform deduplication, which exploits data redundancy and avoids storing duplicated data from multiple users.

A substantial helping of the dataset in big data assignments is redundant. As a result, deduplication knowledge, which eradicates replicas, develops an attractive solution to save disk space and traffic in a big data environment. However, the overhead of extra CPU subtraction (hash indexing) and IO budding introduced by deduplication should be painstaking. Therefore, the net outcome of using deduplication for big data loads needs to be examined.

II. WHY DEDUPLICATION IN BIG DATA?

- More than 2.5 quintillion bytes of data are generated every day. 90% of the total data has been bent just in the past few years alone. To cover such a huge amount of data, storage continues to grow at an volatile rate (52% per year) [1][14]. By the end of 2015, the size of the total formed data will surpass 7.9 zettabytes (ZB). This number is expected to reach 35 ZB in 2020, which has established to be too conservative [2][15].

- Big data assignments have joblessness. On even, 44% of the lively data set in our big data assignments is redundant. Organizing an extra VM yields 97% more redundant data. Using a repetition device in the Hadoop distributed file system (HDFS) presents 19% jobless data on average. The

statistics node covers 25% more jobless data than the name node.

- Additional confusion calculation on the CPU leads to extra power ingesting (around 10%), which results in drive overhead (7%). However, for the benchmarks with a high level of joblessness, the overall energy can be protected (by 43% in our experiment).The deduplication reduces assignment implementation time. There is a strong association between drive influence and the degree of redundancy.

- Deduplication helps assignments utilize more disk amount (3X higher when deduplication is on in our experiment), which leads to, at most, a 45% performance development. Though, due to the above of hash indexing, performance can damage by 161% for some benchmarks in an extreme circumstance.

- In a cross SSD/HDD (Solid State Drive / Hard Disk Drive) environment, deduplication can improve the scheme performance (by up to 17% in our experiment) if the SSD relative is adjusted correctly. However, in a pure SSD situation, deduplication costs presentation and liveliness overhead (about 5% and 6% respectively).

III. LITERATURE SURVEY AND RELATED WORK

Dictionary Reduction model

Traditionally, data reduction has been the result of data compression approaches that use a dictionary model to identify redundancy for short strings (e.g., 16 B), such as the classic LZ77/LZ88 algorithms. Most of these approaches first compute a weak hash of strings and then compare the hash-matched strings byte by byte. Because of their time and space complexity, dictionary model-based compression approaches, such as LZO, LZW, DEFLATE, only compress data in a much smaller region, e.g., data within a file or a group of small files, which trades off processing speed against compression effectiveness.

Delta Compression Technique

The technique is to the indexing issue of delta compression either record the resemblance information for files, instead of data chunks, so that similarity index entries can fit in the memory, or exploit the locality of backup data streams in deduplication-based backup/archiving systems, which avoids the global indexing on the disk

Locality-based approach

Locality-based approaches exploit the inherent locality in a backup stream, which is widely used in state-of-the-art deduplication systems such as DDFS, Sparse Indexing, and ChunkStash. The locality in this context means that the chunks of a backup stream will appear in approximately the same order in each full backup with a high probability. Mining this locality increases the RAM utilization and reduces the accesses to on-disk index, thus alleviating the disk bottleneck.

Similarity based approach

Similarity based approaches are designed to address the problem encountered by locality-based approaches in backup streams that either lack or have very weak locality (e.g., incremental backups). They exploit data similarity instead of locality in a backup stream, and reduce the RAM usage by extracting similar characteristics from the backup stream. A well-known similarity-based approach is Extreme Binning that improves deduplication scalability by exploiting the file similarity to achieve a single on-disk index access for chunk lookup per file.

History Aware Re-writing Algorithm

This existing model is to rewrite duplicate but fragmented chunks during the backup via rewriting algorithm, which is a trade-off between deduplication ratio (the size of the non-deduplicated data divided by that of the deduplicated data) and restore performance. These approaches buffer a small part of the on-going backup stream, and identify the fragmented chunks within the buffer. They fail to accurately identify sparse containers because an out-of-order container seems

sparse in the limited-sized buffer. Hence, most of their rewritten chunks belong to out-of-order containers, which limit their gains in restore performance and garbage collection efficiency.

The Capping algorithm

Capping algorithm are recently proposed rewriting algorithms to address the fragmentation problem. Both of them buffer a small part of the on-going backup stream during a backup, and identify fragmented chunks within the buffer (generally 10-20 MB). For example, Capping divides the backup stream into fixed-sized segments (e.g., 20 MB), and conjectures the fragmentation within each segment. Capping limits the maximum number (say the chunks in the *N of T* containers that hold the least chunks in the segment are rewritten.

3.1 RELATED WORK

Qiang Li, Qinfen Hao, Limin Xiao and Zhoujun Li

[14] proposed VM-base architecture for adaptive management of virtualized resources in cloud computing. The authors also designed a resource controller named Adaptive Manager that dynamically adjusts multiple virtualized resource utilization to achieve application Service Level Objective (SLO) using feedback control theory. Adaptive Manager is a multi-input, multi-output (MIMO) resource controller which controls CPU scheduler, memory manager and I/O manager based on feedback mechanism. For the periodic measurement of the application performance each Virtual Machine has sensor module which transmits information to the adaptive manager. Authors adopted Kernel based Virtual Machine (KVM) as a tool for infrastructure of virtual machine.

Mayank Mishra, Anwesha Das, Purushottam Kulkarni and **Anirudha Sahoo** [15] discussed that live virtual machine migration plays a vital role in dynamic resource management of cloud computing. The authors mainly focused on efficient resource utilization in non-peak periods to minimize wastage of resources. For the

attainment of goals like server consolidation, load balancing and hotspot mitigation, authors discussed three components – when to migrate, which VM to migrate and where to migrate – and approaches followed by different heuristics to apply migration techniques. Authors also discussed virtual machine migration over LAN and WAN with their challenges.

Fan and Bifet [16] pointed out that the terms "big data" and "big data mining" were first presented in 1998, respectively. The big data and big data mining almost appearing at the same time explained that finding something from big data will be one of the major tasks in this research domain. Data mining algorithms for data analysis also play the vital role in the big data analysis, in terms of the computation cost, memory requirement, and accuracy of the end results. In this section, we will give a brief discussion from the perspective of analysis and search algorithms to explain its importance for big data analytics.

Ruijin Zhou, Ming Liu, Tao Li [17] deliberated that the redundancy of big data is measured by three foundations 1) deploying more nodes, 2) increasing the dataset, and 3) using repetition devices. They future about characterize the joblessness of characteristic big data assignments to justify the need for duplication. They deliberated about the analyze and describe the presentation and energy influence brought by repetition under various big data surroundings. In their trials, they identify three bases of redundancy in big data jobs: 1) deploying more nodes, 2) expanding the dataset, and 3) using replication devices.

Min Chen · Shiwen Mao · Yunhao Liu [18] focused on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, they introduce the general background, discuss the technical challenges, and review the latest advances. Finally they examined the several representative applications of big data, including enterprise management, Internet of Things, online social networks, medial applications, collective intelligence, and smart grid.

Dongchul Park, Ziqi Fan, Young Jin Nam, and David H. C. Du [19] discussed on Data duplication (dedupe for short) is a special data compression technique. It has been widely adopted to save backup time as well as storage space, particularly in backup storage systems. Therefore, most dedupe research has primarily focused on improving dedupe write performance. So backup storage dedupe read performance is also a crucial problem for storage recovery.

T.Thamarai Selvan [20] presented a programmatic cloud suitable medical diagnostic computing application which actually gets a pulse modified into 32 bit form data via sensors and plot it as time varying graph. The author drafted an optimal pulse system measurement algorithm and made it venal into the programmatic application which is very efficient in transfiguring data to its core binary format with lose-less nature and plot graph which can be exported into any Big data applications for further analyzing. The algorithm used in this paper can be referenced for creating a resource allocation model in cloud environment running big data applications as it focuses on a schema where the data is processed in real time and processed.

Pritee Patil, Nitin N. Pise [21] discussed on the greatest test for enormous information from a security perspective is the assurance of client's protection. Enormous information as often as possible contains gigantic measures of individual identifiable data and thusly security of clients is a colossal concern. Be that as it may, encoded information present new difficulties for cloud information deduplication, which gets to be significant for huge information stockpiling and preparing in cloud. Customary deduplication plans can't take a shot at encoded information.

Irfan Ahmad Murali Vilayannur Jinyuan Li [22] propose that DEDE, a block-level deduplication system for live cluster file systems that does not require any central coordination, tolerates host failures, and takes advantage of the block layout policies of an existing cluster file system.

Yinjin Fu, Hong Jiang , Nong Xiao[23] proposed that Σ-Dedupe, a scalable inline cluster deduplication framework, as a middleware deployable in cloud data centers, to meet this challenge by exploiting data similarity and locality to optimize cluster deduplication in inter-node and intra-node scenarios.

Lei Wei, Chuan Heng Foh, Bingsheng He and **Jianfei Cai** [24] proposed a heterogeneous resource allocation approach, called skewness-avoidance multi-resource allocation (SAMR), to allocate resource according to diversified requirements on different types of resources. The work includes a VM allocation algorithm to ensure heterogeneous workloads are allocated appropriately to avoid skewed resource utilization in PMs, and a model-based approach to estimate the appropriate number of active PMs to operate SAMR.

Ricardo Koller Raju Rangaswami [25] discussed that e I/O Deduplication, a storage optimization that utilizes content similarity for improving I/O performance by eliminating I/O operations and reducing the mechanical delays during I/O operations. I/O Deduplication consists of three main techniques: content-based caching, dynamic replica retrieval, and selective duplication.

Lauro Beltrao Costa , Samer Al-Kiswany , Raquel Vigolvino Lopes and Matei Ripeanu [26] discussed about the energy trade-offs brought by data deduplication in distributed storage systemsData deduplication enables a trade-off between the energy consumed for additional computation and the energy saved by lower storage and network load.

Shengmei Luo, Guangyan Zhang, Chengwen Wu, Samee U. Khan[27] proposed data deduplication replaces identical regions of data (files or portions of files) with references to data already stored on the disk. Compared with the traditional compression techniques, data deduplication can eliminate not only the data redundancy within a single file, but also the data redundancy among multiple files.

IV. SUMMARY OF RESOURCE ALLOCATION FOR BIG DATA APPLICATION

Table 1 summarizes the work done by various researchers and future work and/or gaps in their existing work.

Table I. Summary of Resource Allocation Techniques

Year	Author	Technique/ Algorithm	Tools used	Future works and/or Gaps in existing technologies	Discussion in terms of Deduplication Workload Resource Allocation for Big Data Application support (can be made by creating / availed by / integrated to / can take this model or algorithm or procedure)
2009	**Qiang Li, Qinfen Hao, Limin Xiao and Zhoujun Li [14]**	Adaptive Management of Visualized resources using Feedback control.	KVM	Only KVM model,I/O Performance, still better modelling can be done	Support can be made by integration with KVM bbut cannot be expected to make load balancing.
2012	**Mayank Mishra, Anwesha Das, Purushottam Kulkarni and Anirudha Sahoo [15]**	Live Virtual Machine Migration	Not mentioned	Only Load on virtual machine is considered. Consumer requirements and priority of the job is not considered.	Support can be made if we create an API which can integrate with the big data application.
2015	**Fan and Bifet [16]**	Data Mining algorithm	Not mentioned	Author has plan to discover more diverse, larger, and faster.	Big Data may be a hype to sell Hadoop based computing systems.
2013	**Ruijin Zhou, Ming Liu, Tao Li[17]**	Advanced hash indexing algorithm	Not mentioned	Writer has plans to modify the benefit of evading jobless disk admissions and the overhead of hashing.	Big data requests counting web hunt, machine learning, analytical inquiry, and categorization
2014	**Min Chen· Shiwen Mao· Yunhao Liu [18]**	Hash algorithm, Analytical Algorithm	Not mentioned	The author has plan to propose a new algorithm on special multimedia event detection using a few positive training	The growth of big data applications accelerates the revolution and innovation of data centers. Many big data applications have developed their unique architectures and directly promote the development of storage, network, and computing technologies related to data center
2016	**Dongchul Park, Ziqi Fan, Young Jin Nam, and David H. C. Du [19]**	Cache Algorithm (LRU)	Not mentioned	In future the author has to extended design outperforms LRU by an average of 64.3%	Big Data dedupe is a specialized technique to eliminate duplicate data so that it retains only one unique data copy on storage
2011	**T.Thamarai Selvan [20]**	Optimal pulse system	Not mentioned	Author has plans to modify the schema to	Schema of this algorithm can be used to test the efficiency

		measurement algorithm		promote more resource allocation optimisation.	of the application with respect to streaming data.
2016	**Pritee Patil, Nitin N. Pise [21]**	Resource allocation through skewness technique	Not mentioned	Extensive performance analysis and test showed that our scheme is secure and efficient under the described security model and very opportune for sizably voluminous data deduplication	Distributed computing is a rising administration display that gives calculation and capacity assets on the Web.
2009	**Irfan Ahmad Murali Vilayannur Jinyuan Li [22]**	Deduplication algorithm	Not mentioned	Author has plan to explore alternate indexing schemes that allow for greater control of deduplication policy	The evaluation of our deduplication techniques using various microbenchmarks and realistic workloads
2012	**Yinjin Fu, Hong Jiang , Nong Xiao[23]**	Data Routing algorithm	Xen Virtual machine	Author has plan achieve in each node by exploiting similarity and locality in backup data streams.	Managing the data deluge under the changes in storage media to meet the SLA requirements becomes an increasingly critical challenge for Big Data protection
2015	**Lei Wei, Chuan Heng Foh, Bingsheng He and Jianfei Cai [27]**	Heterogeneous resource allocation approach, called skewness-avoidance multi-resource allocation (SAMR)	VMware	Workloads are concentrated but the work allocator bases are left after primary allocation table readiness.	The approach can be studied for the work load handling but cannot be considered as a whole as Big data applications must be controlled in more trivial aspects for better and efficient data processing.
2010	**Ricardo Koller Raju Rangaswami [25]**	Replacement algorithm	Not mentioned	The future direction is to optionally coalesce or even eliminate altogether write I/O operations for content that are already duplicated elsewhere on the disk, or alternatively direct such writes to alternate locations in the scratch space	Needs a lot of implementation as the technique proposed has a lot of variation with regard to certain big data data sets.
2011	**Lauro Beltrao Costa , Samer Al-Kiswany , Raquel Vigolvino Lopes‡ and Matei Ripeanu [26]**	Checkpointing is representative for workloads that can benefit from deduplication	Check point application	Author propose an energy consumption model that highlights the same issues and, in spite of its simplicity, can be used to reason about the energy and performance break-even points when	Data deduplication is a method to detect and eliminate similarities in the data.

				configuring a storage system	
2015	**Shengmei Luo, Guangyan Zhang, Chengwen Wu, Samee U. Khan[27]**	Data routing algorithm and Rabin's fingerprint algorithm.	Not mentioned	Author propose that in future we remove the data deduplication ratio in single node with the help of cache container of hot fingerprint based on access frequency	Computing resources of a cluster efficiently and satisfy the applications' demands of parallel processing on big data in the cloud storage system.

V. CONCLUSION AND DISCUSSION

Big data management strategies and best practices are still evolving, but joining the big data movement has become an imperative for companies across a wide variety of industries. Big data analytics is the process of examining large data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits.Many authors have proposed algorithms, methods and techniques for dynamic resource allocation in Deduplication Workload Resource for Big Data Applications environment that support big data applications. We have taken schema from some authors which can be converted to a meaningful insight for our work. In summary, an efficient Resource Allocation Technique should meet following criteria's: Quality of Service (QoS) aware utilization of resources, cost reduction and power reduction / energy reduction.. The ultimate goal of Deduplication Workload Resource Allocation for Big Data Applications is to identify the redundancy of sequences of bytes across very large comparison windows. Sequences of data (over 8 KB long) are compared to the history of other such sequences. The first uniquely stored version of a sequence is referenced rather than stored again.

VI. REFERENCES

[1]. EMC Data Domain. http://www.datadomain.com/.

[2]. IBM ProtecTIER. http://www-03.ibm.com/systems/storage/news/center/ deduplication/index.html.

[3]. Acronis. http://www.acronis.com/backup-recovery/deduplication-roicalculator.html

[4]. B. Efron. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2010.

[5]. U. Fayyad. Big Data Analytics: Applications and Opportunities in On-line Predictive Modeling. http:// big-data-mining.org/keynotes/#fayyad, 2012.

[6]. Dimitrios Zissis and Dimitrios Lekkas, "Addressing cloud computing security issues" in ELSEVIER - Future Generation Computer Systems 28 (2012) 583–592.

[7]. Liang-Jie Zhang, Jia Zhang, Jinan Fiaidhi, J. Morris Chang, "Hot Topics in Cloud Computing" in IEEE Computer Society, ITPro 2012, 1520-9202.

[8]. Robert Grossman, "The Case for Cloud Computing" in IEEE Computer Society, IT Pro 2009.

[9]. B. Efron. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2010.

[10]. U. Fayyad. Big Data Analytics: Applications and Opportunities in On-line Predictive Modeling. http:// big-data-mining.org/keynotes/#fayyad, 2012.

[11]. D. Feldman, M. Schmidt, and C. Sohler. Turning big data into tiny data: Constant-size coresets for k-means, pca and projective clustering. In SODA, 2013.

[12]. J. Gama. Knowledge Discovery from Data Streams. Chapman & Hall/Crc Data Mining and Knowledge Discovery. Taylor & Francis Group, 2010.

[13]. J. Gantz and D. Reinsel. IDC: The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. December 2012.

[14]. Gartner, http://www.gartner.com/it-glossary/bigdata

[15]. Offshore Oil and Gas Supply. Working Document of the National Petroleum Council, 2011

[16]. The Changing Geospatial Landscape. A Report of the National Geospatial Advisory Committee, 2009

[17]. How Big Data Is Changing Astronomy (Again). The Atlantic, 2012

[18]. W. Cox, M. Pruett, T. Benson, S. Chiavacci, and F. Thompson III. Development of Camera Technology for Monitoring Nests. USGS Northern Prairie Wildlife Research Center, 2012

[19]. http://www.groundcontrol.com/Oil-And-Gas_Satellite.htm.

[20]. Bhagwat D, Eshghi K, Long D D E, et al. Extreme binning: Scalable, parallel deduplication for chunk-based file backupCModeling, Analysis & Simulation of Computer and Telecommunication Systems, 2009. MASCOTS'09. IEEE International Symposium on. IEEE, 2009: 1-9.

[21]. Dong W, Douglis F, Li K, et al. Tradeoffs in Scalable Data Routing for Deduplication ClustersCFAST. 2011: 15-29.

[22]. You L L, Pollack K T, Long D D E. Deep Store: An archival storage system architectureCData Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on. IEEE, 2005: 804-815.

[23]. Eshghi K, Tang H K. A framework for analyzing and improving contentbased chunking

[24]. Liu C, Lu Y, Shi C, et al. ADMAD: Application-Driven Metadata Aware De-duplication Archival Storage SystemCStorage Network Architecture and Parallel I/Os, 2008. SNAPI'08. Fifth IEEE International Workshop on. IEEE, 2008: 29-35.

[25]. Bobbarjung D R, Jagannathan S, Dubnicki C. Improving duplicate elimination in storage systemsJ. ACM Transactions on Storage (TOS), 2006, 2(4): 424-448.

[26]. Kruus E, Ungureanu C, Dubnicki C. Bimodal content defined chunking for backup streams CProc of the USENIX FAST10, Brekeley, CA:USENIX, 2010: 239-252

[27]. Zhu B, Li K, Patterson R H. Avoiding the Disk Bottleneck in the Data Domain Deduplication File SystemCFast. 2008, 8: 1-14. 29Bloom B H. Space/time trade-offs in hash coding with allowable errorsJ. Communications of the ACM, 1970, 13(7): 422-426. 30Lillibridge M, Eshghi K, Bhagwat D, et al. Sparse Indexing: Large Scale, Inline Deduplication Using Sampling and LocalityCFast. 2009, 9: 111- 123.

[28]. Broder A Z. On the resemblance and containment of documentsC//Compression and Complexity of Sequences 1997. Proceedings. IEEE, 1997: 21-29.

[29]. Debnath B, Sengupta S, Li J. ChunkStash: speeding up inline storage deduplication using flash memoryCProceedings of the 2010 USENIX conference on USENIX annual technical conference. USENIX Association, 2010: 16-16.

An Improved Real Unidentified Protected Routing (Rupr) For Manets in Wireless Networks

S. Shahul Hammed
Assistant Professor, Department of Department of Computer Science and Engineering, Karpagam Academy of Higher Education, Coimbatore, India
Email: shahul.y2s@gmail.com

ABSTRACT

The mobile ad hoc networks (MANETs) are wireless and heterogenic network topology middling, which might go through from much security criticism. The key factor of networks is to transmit the packet in safe manner from source to terminal nodes in adversarial surroundings such wireless node message have three factors are the (i) mobile traffic, (ii) node attack and (iii) packet accessing of in-between nodes. The existing protocol mechanism is the source of verification grouping name, and safe routing procedure. The proposed system presents a conviction (trust) based route finding protocol technique is established real unidentified protected Routing (RUPR) with Trust based model. An Improved RUPR protocol idea is to protect the adjacent nodes attack by the way of encryption and decryption in path-request and path-reply. With the help of the trust based model (QoS routing protocol), the will be more dynamic in recognizing connection (path) failures, caused either by the mobility or opponent attacks. The scheming trust assessment of the in-between node in MANET routing can helps to circumvent the multipath message transfer interruption between nodes.

Keywords : Mobile ad hoc networks, Heuristic Load balancing, dynamic channel allocation, Graph

I. INTRODUCTION

Mobile ad hoc networks (MANETs) are susceptible to protection threats unpaid to the natural individuality of such networks, such as the open wireless average and dynamic topology. It is hard to provide trusted and secure connections in adversarial locations. On one hand, the adversary's external a network may deduce the in order about the communicating mobile nodes or transfer flows by passive transfer examination, still if the messages are encrypted. On the additional, the mobile nodes contained by the network cannot be constantly expectation, since a appropriate mobile nodes may be listed by opponent and become malicious. As a result, unidentified messages are important for MANETs in adversarial positions, in which the nodes classifications and links are restored by random numbers or pseudonym for protection reason.

A mobile Ad hoc network (MANET) is a group of independent mobile nodes proficient of message with both of via network paths. Mobile nodes in a framework have incomplete broadcast range; message is attained by creation use of nodes to ahead packets to other nodes, which thereby have to control as routers. Searching a pathway between two message end positions in an ad hoc network is non-slight: node mobility consequences in extremely vibrant network topologies. These types of networks are rapidly arranged, as they don't need any communications in place. MANETs are highly attractive in a selection of circumstances: tragedy recovery-where the whole communication locations might have been shattered, business meetings- where a assembly of community have to split resources and message with each other, communication over rough territory – where creating a infrastructure is not price effective. Ad hoc networks can also be used to organize multimedia

locations; though capable routing protocols have to be residential before this can be realized.

The rest of this paper is organized as follows. In Section 2 review the Literature survey. The proposed models and descriptions are described in Section 3. Finally conclude the paper in Section 4.

II. Literature Survey

In [1] authors proposed to protect confidentiality over a message network, a excess of unidentified protocols have been illustrated along with several experiential examinations into explicit opponent attacks over those networks. However, no recognized classification exist that tackle secrecy in the various position of together wired and wireless unidentified message networks. The corresponding model provides a new novel classification which discovers the 3 key methods of secrecy property, opponent ability, and network type. In [2] authors addressed the rapid version to dynamic path constraints, low dispensation and memory transparency, low network consumption, and establishes single transmit paths to terminals within the ad hoc network. It uses destination progression numbers to provide loop autonomy at all times (even in the face of irregular delivery of routing direct messages), keep away from problems (such as "including to infinity") connected with traditional space vector protocols. In [3] presented the DSR permits the network to be completely personality organizing and personality configuring, without requires for any previous network communications or management. The protocol is designed of the two major devices of "Path Learning" and "Path Preservation", which effort jointly to permit nodes to learn and preserve routes to random terminals in the ad hoc network. In [4] authors discussed the ANODR, an anonymous on-demand routing protocol for networks organized in aggressive surroundings. The authors discussed the two strongly connected problems: For path secrecy, ANODR checks well-built opponents from tracing a message flow reverse to its source or terminal; for position retreat, ANODR guarantees that opponents cannot learn the actual uniqueness of local transmitters. In [5] authors propose the security and privacy in mobile ad hoc networks has been an important subject over the last few years. Existing studies work has so far alert on provided that security for path and message substance, but nothing has been

complete in hold to given that confidentiality and secrecy over these networks. Authors discussed a optimal distributed routing protocol which assurances protection, secrecy and elevated dependability of the established path in a aggressive location, such as an ad hoc wireless network, by encrypting the steering message description and abstention from using untrustworthy middle nodes. In [6] authors proposed the safety, secrecy, and scalability are silent significant subjects for mobile ad hoc network steering protocols. To representation the restrictions of some accessible mobile ad hoc network steering protocols with safety and secrecy condition and examine their scalabilities. Based on the examination the new unidentified dynamic source routing protocol to offer three level of safety protection.

III. Proposed Methodology

A. Group Signature

The group Signature is a technique is for allowing users of a group to notice secretly in a MANET routing protocol. Group Signatures can be analyzed as conventional public key names with further privacy features. This approach is to execute a collection key conformity protocol at the opening of each time period and use the resultant group key as the frequent initialization and scalable. The more proficient method is to utilize a group type agreement protocol in organize to consent on the common limitation and group manager to produce and allocate this opening value. This system has group manager, who is reply for adding new users and repealing signature of personality nodes in secrecy are given to a group manager.

B. Routing Design

In rouging design process the source node establish its link to transmit the HAI or HELLO message to the terminal mobile nodes. It verifies the communication link and builds clear by way of broadcasting request to the terminal node.

Source node: A primarily sends the message to the terminal node E with assembly key K and make encryption when the message is broadcast (1).

$$S \rightarrow [P_A^-, G_A^+, K_{AE}, O_A^-] \qquad (1)$$

Intermediate node: The middle of the node which collects the packets from source node A and the further encryption, before transmit the message to the End node E (2).

$$I \rightarrow [P_B^-, G_B^-, O_B^-] \qquad (2)$$

Terminal node: Terminal node *E* receives the message from middle of the node, which uses public key admission to the undisclosed message. The node *E* is prepared to path reply after accept the packet and reply to node *A*.

When forwarding a message in each middle of node is dependable for verifying that the message is properly acknowledged by the subsequently node, however appropriate to the dynamic topology and the conditions of the wireless networks it may arise some circumstances where a node doesn't receive the acknowledgement of response from link layer of a given message, consequently it retransmits the similar message it until attains a threshold value of attempts. At Every time the number of efforts was accomplished the corresponding node reflect on this link as broken than it removes each path surrounding this link from its cache than it produces a path error message to report to the source node and all intermediate nodes about this path failure in the similar way at each middle node removes all routes containing this route until the path error packet appears to its terminal which decides to initiate a new path request or to search a new path in its cache.

C. Trust Based Routing Protocol

The trust based protocol is the authenticated as the dimension of individual certainty about the mobile activities of an enough entity. It is the likelihood through which an individual node routine of unidentified routing in adversarial location. Trust node is linked to routine of mobile nodes in the packet reputation and reference. The node of anonymous routing is in related environment response for minimizing the delay in message transmission. Trust in MANETs is a degree of the idea that a node in a network or mediator in a distributed method will transmit out tasks. In this path observation, observer approximations the trust of his single-hop neighbour related on its own evaluation. Therefore, the trust value (Tr) is the probability of a subjective chance that a trust or uses to choose whether or not a trustee is

dependable. In the shortest observation, to assume the both observer can eavesdrop packets ahead by an observed node and evaluate them with unique packets, so that the viewer can recognize the malicious behaviours of the practical node. Therefore, the listener node can analyse the confidence values of its neighbours. In order to attain less unfairness trust value (UTr), it also considers other observers estimations in this model. If the trust value (Tr) is fewer than the predefined entry value (λ), the mobile node will be predetermined as un-trusted mobile node and will not be measured for additional transmission.

D. *Improved Real Unidentified Protected Routing (RUPR) Protocol*

The improved real unidentified protected routing (RUPR) with Quality of Service (QoS) parameters are searching a different path between starting and ending node. The multi-path denotes a link between neighboring nodes that may divide up and optionally join up. This should not be incorrect for multicast fading, which will be noticed. Depending on the QoS metrics in use, it is also possible to divide a QoS conditions into dissimilar sub-conditions. The multi-path routing exists no link between nodes S and T that can verify a bandwidth constraint so the conditions and the links are dividing up at neighboring node. As both paths meet at neighbor node, they join again. The goal of unidentified Secure QOS Routing is to decrease the message transparency when building a multicast tree by controlling between single path routing and multi-path routing. When a mobile nodes n needs to join a previous multicast hierarchy, a single path to the hierarchy score is searched using a single transmit searching algorithm. During this path learning process, the QoS conditions are checked at middle of the node. Regard as two middle nodes a and b with an individual part of the previously learned gateway. If b is the subsequently node selected by the single path algorithm, but the link (a, b) breaks the QoS conditions, then instead of the throw messages to its other neighboring mobile nodes to divide up the search process. If more than one feasible alternate path is detected, a choose the best path. The number of divide ups can be controlled by identifying a maximum division level.

IV. Performance Evaluation

Packet delivery Ratio (PDR):

The ratio of the data packets delivered to the destinations to those generated by the Constant Bit Rate (CBR) sources. The PDR shows how successful a protocol performs delivering packets from source to destination in figure 1. The higher for the value give use the better results. This metric characterizes both the completeness and correctness of the routing protocol also reliability of routing protocol by giving its effectiveness.

PDR is the ratio of the number of data packets received by the destination node to the number of data packets sent by the source mobile node. It can be evaluated in terms of percentage (%). This parameter is also called "success rate of the protocols", and is described as follows:

$$PDR = \left(\frac{SendPacketno}{Receivepacketno}\right) \times 100 \quad (3)$$
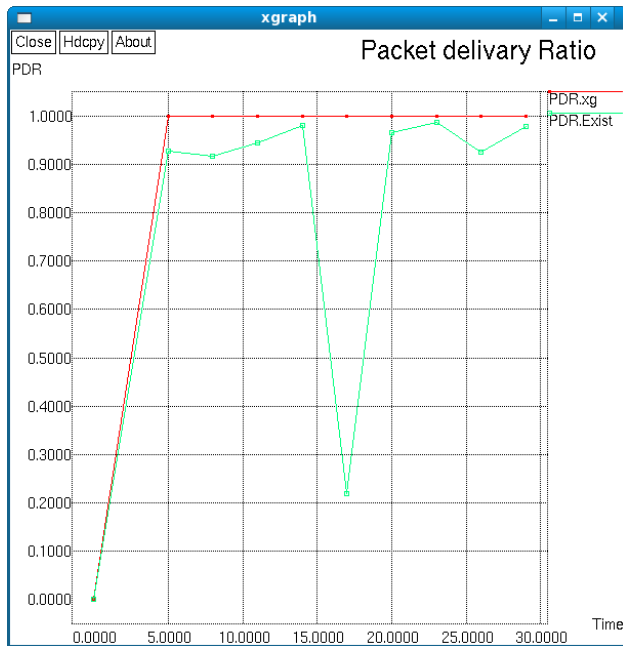
Throughput:

 The ratio of the total amount of data that reaches a receiver from a sender to the time it takes for the receiver to get the last packet is referred to as throughput. It is expressed in bits per second or packets per second. Factors that affect throughput include frequent topology changes, unreliable communication, limited bandwidth and limited energy. A high throughput network is desirable. Throughput is the average rate of successful message delivery over a communication channel. This data may be delivered over a physical or logical link, or pass through a certain network node.

$$X = \frac{C}{T}(4)$$

Where X is the throughput, C is the number of requests that are accomplished by the system, and T denotes the total time of system observation.
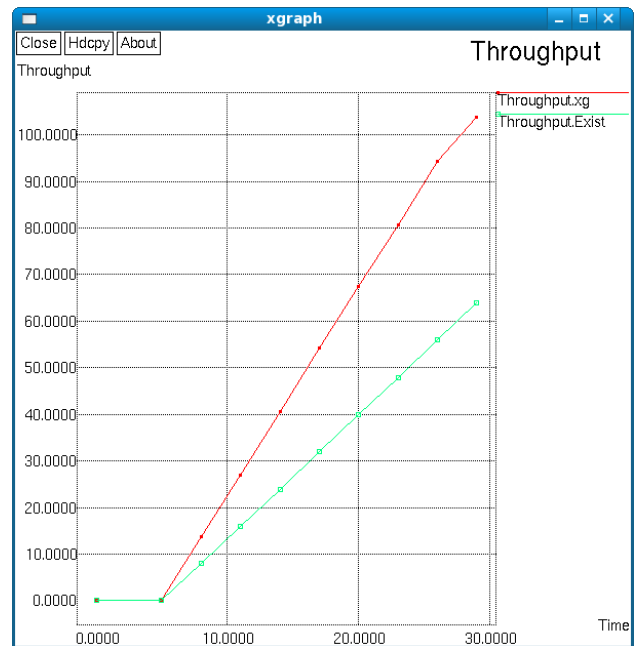
V. Conclusion and Future Work

In this paper proposed the Real Unidentified Protected Routing (RUPR) Protocol mechanisms for protected routing for Manets in adversarial environment. The proposed technique using a general approach outperforms the uses dependence values to support message ahead by preserving a trust offset for each node. If the trust value falls lower than a threshold, the subsequent middle node is malicious node. In this proposed system, certified node has high energy and message delivery ratio can be enhanced extensively with lessening common end to end delay by increasing trust value.

In future work, we intend to improve the proposed algorithm to develop the experimental methods for unidentified protected protocol optimization to control the attacks of the result data.

VI. References

[1] D. Kelly, R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "Towards a taxonomy of wired and wireless anonymous Networks," in Proc. IEEE WCNC'09, Apr. 2009.

[2] C. Perkins, E. Belding-Royer, S. Das, et al., "RFC 3561 - Ad hoc On- Demand Distance Vector (AODV) Routing," Internet RFCs, 2003.

[3] D. Johnson, Y. Hu, and D. Maltz, "RFC 4728 - The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," Internet RFCs, 2007.

[4] J. Kong and X. Hong, "ANODR: ANonymous On Demand Routing with Untraceable Routes for Mobile Ad hoc networks," in Proc. ACM MobiHoc'03, Jun. 2003, pp. 291–302.

[5] J. Kong, X. Hong, and M. Gerla, "ANODR: An identity-free and ondemand routing scheme against anonymity threats in mobile ad hoc networks," IEEE Trans. on Mobile Computing, vol. 6, no. 8, pp. 888–902, Aug. 2007.

[6] A. Boukerche, K. El-Khatib, L. Xu, and L. Korba, "SDAR: a Secure Distributed Anonymous Routing Protocol for Wireless and Mobile Ad hoc Networks," in Proc. IEEE Int'l Conf. Local Computer Networks (LCN'04), Nov. 2004, pp. 618–624.

Multi Biometric Security Pattern Using finger vein and finger print

S. Ilankumaran[1], Dr. C. Deisy[2]
Department of Information Technology, K.L.N. College of Engineering, pottapalayam, Sivagangai, Tamil Nadu, India
ilankumaran.s@klnce.edu
Department of Computer Science, Thiagarajar College of Engineering, Madurai, India
cdcse@tce.edu

ABSTRACT

Today India is moving towards cashless transactions , transformation its cities into smart cities , constructing varies potential nuclear power plant , providing aadhaar card for each citizen in the country which leads us to the need of a promising technology for the above all functionality , which should provide better security for all endorsement system because of above all functions should be performed only be the authenticated person otherwise its effect will be enormous by affecting individual person or whole country. For the promising technology , here we have proposed a new methodology which use finger vein and finger print as a key factor in which we extract both the finger vein and finger print from user to find out whether they are the authenticated users before they access the all endorsement system. For the feature extraction, Competitive Matching Code (CMC) has been used. This identification system will be well suited for endorsement system comparing to other biometric system and provide better results on EER and total processing time

Keywords: Competitive Matching Code, EER, Nearest Neighbour

I. INTRODUCTION

Biometric Authentication system is a process that validates the individual identity of person who is going to access the potential systems by evaluating their biological and behavioural traitssuch as it can besignature, palm prints, fingerprints, iris patterns and voice spectrum of that particular person. But basically there some authentication schemes are used in real world to examine whether person is authenticated user, first schemes which make use of external hardware material such as the Id cards, metal keys to authenticate them. Some schemes is based on secret software security key such that password, pattern, pin etc. finally some scheme is evaluating the biometric recognition of a particular person before they access the potential system. But First scheme can be easily hacked by duplicated the external hardware material, second scheme lead

authenticated user not able to access in case if they forget their secret password lead to reset the entire security system. This problems can be come over by using Biometric authentication system in which the user need not remember any secret keys and user should physically present at the time of authenticating while using the biometric system such as finger print, palm prints, iris patterns. but it can easily extracted from the authenticated user without any acknowledge of them. it became a major problem in biometric authentication system it can resolve by the usage of finger vein which is unique identification declared in 2008 which cannot extract without any acknowledge of particular user and which overcome all the above problem. We consider the finger vein and finger print as key factor for the authentication system.

II. SYSTEM OVERVIEW

2.1. Capturing Finger vein image.

Finger vein images are captured by using near infra-red (NIR) camera through light transmission methodologies which are used here. In the method the finger dorsal is placed at NIR and finger vein is captured by passing the light through the finger. To get high contrast images, we have adapted light transmission method as shown in fig1
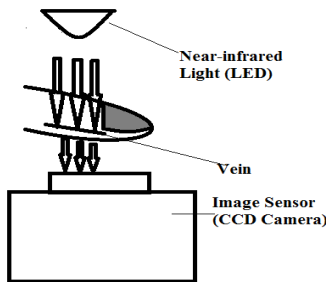
Figure 1: Light Transmission method

Figure 2: (a) Finger vein Image as raw

Figure 2: (b) Finger Print Image as raw

2.2. Capturing finger print Image

Finger print can be captured using optical scanner and capacitive scanner .Optical scanner is basically works as digital photograph, initially capture the photograph of finger and process it through ccd sensor and cmos sensor to avoid dirty black photograph and finally

produce the image from which the finger print image is obtain. Capacitive scanner electrically measuring the distance of the hollows between the ridges in the particular finger and build up the finger print image based on the measuring distance.
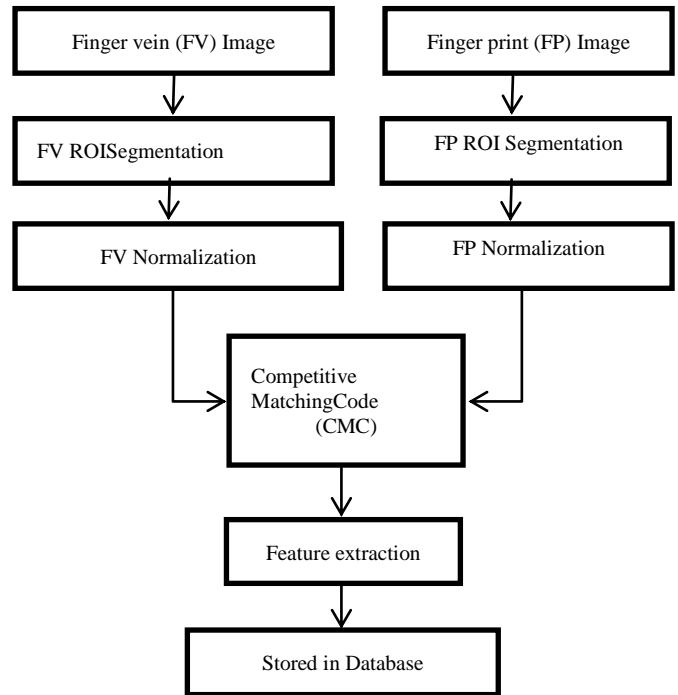
Figure 3: Proposed Verification frame work

2.3. Recognition Frame work:

The proposed recognition frame work is shown in Figure 3. After capturing both the images, the pre-processing model fulfils the ROI extraction. The extracted ROI of the finger vein image is normalized for size and intensity. Similarly the finger print image is pre-processed and the normalized. Subsequently, a feature extraction algorithm called which used on the normalized sub-image. Gabor filters are applied to these normalised finger vein and finger print images and their respective output are used to obtain the magnitude-preserved CMC code [1].CMC code is a state of biometric algorithm known as a competitive code which preserves the magnitude information of the filtered image and is used for further processing. The extracted features corresponding to the code map is stored in the Nearest Neighbour (NN) classifier and used for the verification process.

2.4 Finger print Segmentation

We have adapted the well know traditional algorithm which is used for segmentation of oldest biometric security factor which is fingerprint by mean and variance based algorithm defined by Jun ma et.al [10] In generally finger print pattern composed by two regions they are foreground and background .among these two, foreground consists of important information which is essential thing for recognising correct finger print pattern in biometric authentication system .Since background consists of noise regions here it contributes to the extraction of trivial details in the process .To restricts those type of trivial details, we perform segmentation for finger print. The main purpose of segmentation is to separate the foreground and background

Mean and variance based algorithm

1) Start with Divide the fingerprint image I into non over lapping blocks with size M*N.
2) Compute the mean value Mean for each Block

$$\text{Mean}= \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} I(i,j)}{M*N}$$

Where I (i, j) is the pixel gray value of the ith row and the jth column.

3) Use the mean value to drive the variance value for each block

$$\text{Variance} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[(i,j)-mean]^2}{M*N}$$

4) Select a threshold value which works on different images. If the Variance is larger than threshold value then block is considered as foreground otherwise that is background.

2.4.1. Finger vein Segmentation

The fingers are located in the middle of the raw images. In the finger vein images, the horizontal direction on the finger edges has obvious jumps. Thus, we used a gradient operator as shown in Figure 4 to find the vertical lines which are the finger edges. In moving the mask from the middle of the image to the side, an edge

was detected if the norm of the gradient was higher than the threshold. The left edge was obtained by image convolution by using the gradient operator shown in Figure 4(a), and the mask in Figure 4(b) is to detect the right one. The threshold was experimentally set as 212.

1	0	-1
2	0	-2
3	0	-3
4	0	-4
5	0	-5
6	0	-6
7	0	-7
6	0	-6
5	0	-5
4	0	-4
3	0	-3
2	0	-2
1	0	-1

-1	0	1
-2	0	2
-3	0	3
-4	0	4
-5	0	5
-6	0	6
-7	0	7
-6	0	6
-5	0	5
-4	0	4
-3	0	3
-2	0	2
-1	0	1

Figure 4 (a)   Figure 4 (b)

Figure 4: Edge Detection Masks

The edge lines in the X (horizontal) direction were set to 100 pixels inwards the lower and upper bounds of the captured finger area. Then, the rotational alignment algorithm proposed in [4] was applied. We normalized every row in the finger area to the same length by using linear interpolation. The final size of the finger regions was normalized to 100 x200 pixels. The segmented finger vein image is shown in Figure 5.
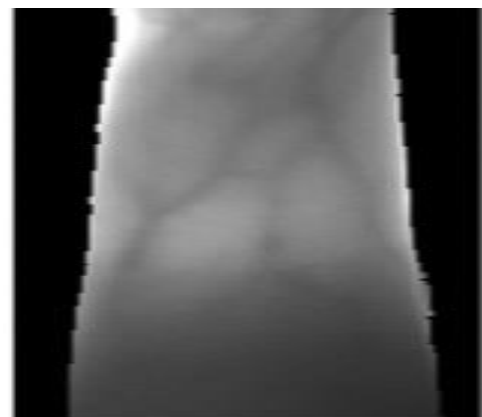
Figure 4: Finger vein image after Segmentation

Figure 5: finger print image after Segmentation

2.5. Feature extraction

The sub images of the finger vein and finger print are applied to the Gabor filter. A best extraction tool Gabor filters which have been widely used in biometric system gives three types of output such as magnitude code, Phase code and Orientation code. These outputs can be used both by individual or combining all the codes in order to apply to a system. A **CMC** Coding scheme that combines the magnitude code and orientation code of Gabor filter [1] is adapted in this paper. A two dimensional even-symmetric Gabor filter can be defined in equation (4)

$$G(x, y) = e^{(-\frac{x_\theta^2}{\sigma_x^2}+\frac{y_\theta^2}{\sigma_y^2})} \cos(2\pi f x_\theta) \quad (4)$$

$$\text{Where} \begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (5)$$

Here, f denotes the filter centre frequency is the orientation parameter as in equation (5), and

$$\theta = \frac{j\pi}{J} \qquad , J = \{0,1,2....,J-1\} (6)$$

σ_x and σ_y, are the standard deviations of the elliptical Gaussian envelope, x_θ and y_θ are rotated versions of the coordinates (x, y) of the Gabor filter. The number of Gabor filters J is fixed to 6.

2.6. Magnitude Preserved CMC Scheme

The **CMC** code scheme [1] has been widely used in palm-print and FKP recognition. The **CMC** coding scheme that combines both the finger vein and finger print images is given in Algorithm 1. The Gabor filter is applied to finger vein and finger print images and the orientation code for finger vein and finger print images are OV Code, OPCode respectively and the Magnitude code for finger vein and finger print images are MVCode, MPCode respectively.

Algorithm 1:

Comparative Competitive Coding scheme

Input OVCode, MVCode, OPCode, MPCode

Output CMCcode

For all CMC codeDo

    If OVCode(x,y)= OPCode(x,y)then

        CMC code=OVcode(x,y)

    Else if MVCode (x, y) <MPCode (x, y) then

        CMC code (x, y) = MVCode (x, y)

    Else if MVCode (x, y) >MPCode (x, y) then

        CMC code(x, y) = MPCode (x, y)

    Else if MVCode(x, y) = MPCode (x, y) then

        CMC code(x, y) = J

In Algorithm 1, the outputs of the Gabor filter, i.e. Magnitude code and Orientation code are considered as the input.
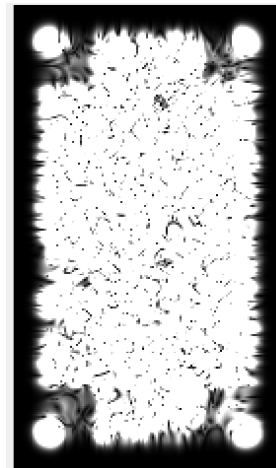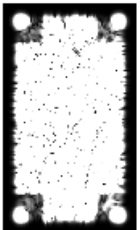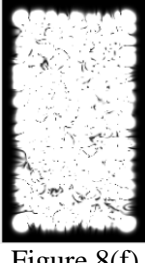
Figure 6: Final image stored in database

The Gabor filter results for some of the pixel will not have much variation, due to low quality image. So the lower value of Magcode that corresponds to the Oricode value is stored. The resulting output iscalled as CMC code. The final image stored in the database is given in Figure 6.

III. Results and Discussions

Pre-processing of images is done for ROI extraction.The segmentation of finger print is done by mean and variance. The edges of the finger vein image are detected using edge detection masks shown in Figure 5 as explained in section 2.4 with a threshold of 212.The ROI of finger vein are extracted .The ROI extracted for sample images of the person, are given in Fig.7.

	Sample 1
Finger vein after Registration and ROI extraction.	
Finger print after Registration and ROI extraction.	

	Sample 1	
	Oricode	MagCode
Finger vein image	Figure 8(a)	Figure 8(b)
Finger print image	Figure 8(e)	Figure 8(f)
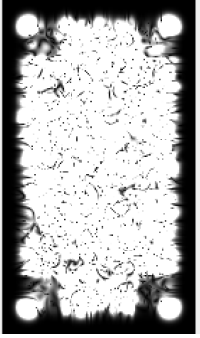
CMC code	Figure 8(i)

Figure 8: CMC Code formulations.

IV. Conclusion and Feature work

This paper presents a new methodology to improve the performance of authentication system using finger vein and finger print. The EER is 1.5441 % where the threshold level is 0.256.0. The advantage of using this method is finger veins and finger prints are the biometric feature which seems to be too tough to get from the user without their knowledge, comparing to the other biometric features. So this method will provide better choice for authentication system, the performance and accuracy can be increased using the proposed feature work.

V. REFERENCES

[1] Wenming Yang, XiaolaHuang, "Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion "Information Sciences 268 (2014) 20–32.

[2] Ajay Kumar, K. VenkataPrathyusha, "Personal Authentication Using Hand Vein Triangulation and Knuckle Shape", IEEE Transactions on image processing, vol. 18, no. 9, sept.2009.

[3] L. Zhang, L. Zhang, D. Zhang, H. Zhu, "Online finger-knuckle-print verification for personal authentication", Pattern Recognition 43 (7) (2010) pp.2560–2571.

[4] Zhi Liu and Shangling Song, "An Embedded Real-Time Finger-Vein Recognition System for Mobile Devices", IEEE Transactions on Consumer Electronics, Vol. 58, No. 2, May 2012

[5] J. Yang, X. Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification", Pattern Recognition Letters 33 (5) (2012) pp.623–628.

[6] W.M. Yang, Q. Rao, Q.M. Liao, "Personal identification for single sample using finger vein location and direction coding", 2011 International Conference on Hand-based Biometrics (ICHB), 2011, pp.1–6.

[7] W.K. Kong, D. Zhang, "Competitive coding scheme for palm print verification", 16th International Conference on Pattern Recognition (ICPR), 2004, pp. 520–523.

[8] V.Ramya, P.Vijaya Kumar, B.Palaniappan, "A Novel Design of Finger Vein Recognition for Personal Authentication and Vehicle Security", Journal of Theoretical and Applied Information Technology 10th July 2014. Vol. 65 No.1.

[9] Jose Anand, T. G. Arul Flora, Anu Susan Philip, "Finger-vein based biometric security system", IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308, 2013.

[10] Jain, A. K., Flynn, P. J. & Ross, A. eds., 2007. Handbook of biometrics. Springer.

[11] Jun Ma, Xiaojun Jing, Yuanyuan Zhang, Songlin Sun, Hai Huang "Simple effective finger print segmentation algorithm for low quality images"

Analysis and Non-linear Control of sepicdc-dc converter

Abarna. R[*1], Vairaprakash. P[2]

*1 PG Student, Department of Electrical and Electronics Engineering, Thiagarajar college of Engineering, Madurai, Tamil Nadu, India
abarna.rjkmr@gmail.com

*2 Assistant Professor, Department of Electrical and Electronics Engineering, Thiagarajar college of Engineering, Madurai, Tamil Nadu, India
vairaprakashbe@gmail.com

ABSTRACT

SEPIC converter has various applications as a power conditioning system because it has a non-pulsating input current and its voltage can be either stepped-down and stepped-up. However, due to the non-linear characteristics of a this type converter, non-linear control techniques are more suitable to attain better performance in voltage regulation. Due to load and output voltage variations, good performance can be achieved through the Non-linear Control because SEPIC converter is a Non-linear system. The performance of the developed controllers with the SEPIC converter is performed through Matlab environment. The proposed system is suitable for all real-world commercial applications, like the power supplies for medical equipment, computer power supplies, uninterrupted power supplies, etc. Various control techniques such as PI controller, fuzzy logic controller, sliding mode controller are used for analysis using Matlab and results are made to compare. The results are compared for change in load variations

Keywords : Converter, Dynamic Performance, Voltage Regulation, Non-Linear Control, Sepic, Smc.

Abbreviations and Acronyms : CCM-Continuous Conduction Mode; VRM-Voltage Regulator Module; SMC-Sliding Mode Controller; FLC-Fuzzy Logic Controller; D-Dutycycle; SEPIC-Single Ended Primary Inductor Converter; MOSFET-Metal Oxide Semiconductor Field Effect Transistor

I. INTRODUCTION

A DC-DC type of converter is a small, light, and highly efficient DC power supply using semiconductor switching elements. DC to DC converters are generally used to convert an input DC voltage to an output DC voltage. Such converters may step down or step up the input DC voltage. Due to their ability to receive low input voltages and consume low power, DC-DC converters have been widely used in all types of electronic products. A DC-to-DC converter provides the functions of regulating the voltage level from a DC input voltage, such as boost or buck voltage conversion, and of maintaining the regulated voltage at the desired level. DC to DC converters are widely used for battery-powered electronic equipment, renewable energy systems, and voltage regulator modules (VRM) to produce a regulated voltage or current derived from an unregulated power supply. This converter is called a switch-mode power supply. The basic converter topologies such as CUK and SEPIC and etc., converters can be implemented for this analysis. The project is proposed with sepic converter which is used to improve the dynamic performance, using various control techniques such as PI controller, fuzzy logic controller and sliding mode controller.

The most commonly used converter is a boost converter. Boost converters are step-up converters in which output a voltage higher than the voltage that is input to the converter. The standard boost converter has an output that is equivalent to the input voltage divided by the duty cycle.

$$V_{out} = \frac{V_{in}}{(1-D)}$$

$$.....(1)$$

A sliding mode controller for the SEPIC is proposed to ensure the stability under any operating conditions, and

to have a better static and dynamic performances for any change in its input as well as load changes and component variations. In sliding mode control, a state trajectory moves back and forth around a certain average surface in the state space. Four sliding variables are considered in this paper since the SEPIC is a fourth order system. The SMC control technique offers several advantages such as stability even for a large supply and load variations, robustness, good dynamic response and simple implementation. Its capabilities result in improved.

II. SEPIC CONVERTER

A SEPIC is a DC-DC converter which has the ability to convert input voltage to an output voltage which can be stepped up or stepped down . The switch of the SEPIC is controlled by changing the duty cycle. This enables close and open conditions. A SEPIC is like a buck-boost converter. However, it has the unique feature of giving a non-inverted output. This means that the output is always the same polarity as the input. A series capacitor is used to couple the energy from the input to the output. The SEPIC can respond quickly to a short-circuit condition, and it works as a true shutdown mode when the switch is turned off and its output drops to 0V following a fairly hefty transient dump of charge.

The SEPIC is useful in applications in which the voltage can be above or below that of the regulator's intended output. The SEPIC transfers energy through the switching operation between the capacitors and the inductors. This is done in order to convert from one voltage to another. The amount of energy is controlled by switch S, which is a transistor such as a MOSFET, IGBT, etc. MOSFET offer a much higher input impedance and a lower voltage stress and do not require biasing resistors. In addition, MOSFET switching can be controlled by differences in voltage rather than a current.

III. MODES OF OPERATION
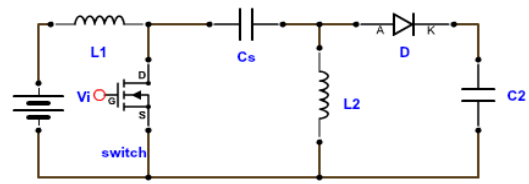
Figure 1. Circuit diagram of sepic converter

Figure 1.shows the circuit diagram of single ended primary inductor capacitor. Single ended primary inductor converter (SEPIC) is a type of converter that allows the electrical potential i.e. voltage at its output to be greater than or less than to that at its input. The output of the SEPIC converter is controlled by the duty cycle of the switch. The SEPIC converter exchanges energy between the capacitors and inductors in order to convert from one voltage to another. The amount of energy exchanged is controlled by switch S. The energy to increase the current in inductor L1 is coming from the input source.

A SEPIC is said to be in continuous-conduction mode if the current through the inductor L_1 never go down to zero. During a SEPIC's steady-state operation, the average voltage across capacitor C_S (V_{C_S}) is equal to the input voltage (V_{IN}). Because capacitor C_S blocks direct current, the average current across it (I_{C_S}) is zero, making inductor L_3 the only source of load current. Hence the average current through inductor L_3 is the same as the average load current and hence independent of the input voltage. Looking at average voltages, the following can be written:

$$V_{IN} = V_{L1} + V_{Cs} + V_{L2}$$
….(2)

Because the average voltage of V_{Cs} is equal to V_{IN}

$$V_{L1} = -V_{L2}.$$
….(3)

For this reason, the two inductors can be wound on the same core. Since the voltages are the equal in

magnitude, their mutual inductance effect will be zero. Here it is assumed that the polarity of the coil is correct. As the voltages are the equal in magnitude, the ripple currents of the two inductors will be equal in magnitude.
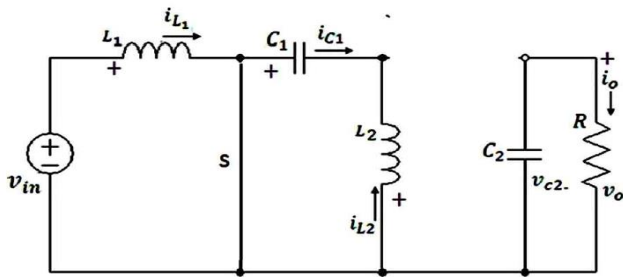
Figure 2. Switch ON condition

Figure 2.shows the working condition of sepic converter during ON state. When switch is turned on, current I_{L1} increases and the current I_{L3}increases in the negative direction. The energy to increase the current I_{L1} comes from the input source. Since Q_1 is a short while closed, and the instantaneous voltage V_{Cs} is approximately V_{IN}, the voltage $V_{L3}$is approximately $-V_{IN}$. Therefore, the capacitor Cs supplies the energy to increase the magnitude of the current in $I_{L3}$and thus increase the energy stored in L3.
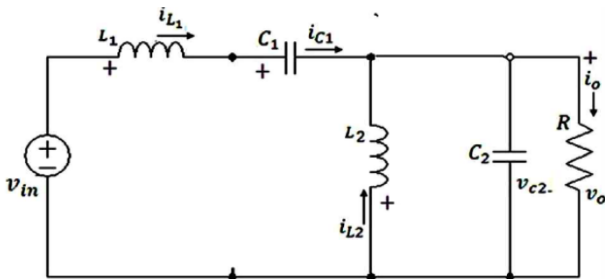
Figure 3. Switch OFF condition

Figure 3.shows the working condition of sepic converter during OFF state. When switch (Q_1) is turned off, the current I_{Cs} becomes the same as the current I_{L1}, as the inductors will not allow instantaneous changes in current. Current I_{L3}will continue in the negative direction, in fact it never reverse direction. It can be seen from the diagram that a negative I_{L3}will add to the current I_{L1} to increase the current delivered to the load.

By Using Kirchhoff's Current Law

$I_{D1} = I_{Cs} - I_{L2}.$

$$.....(4)$$

So while switch is off, power is delivered to the load from both L3and L1. Coupling capacitor (C_S), is charged by L1 during this off cycle, and will recharge L3 during the on cycle. The boost/buck capabilities of the SEPIC are possible because of capacitor C_S and inductor L3. Inductor L1 and switch Q_1 create a standard boost converter, which generates a voltage (V_{Q1}) that is higher than V_{IN}. Its magnitude is determined by the duty cycle of the switch Q_1. Since the average voltage across C_S is V_{IN}, the output voltage (V_{OUT}) is

$V_{OUT} = V_{Q1} - V_{IN}.$

$$.....(5)$$

If V_{Q1} is less than double of V_{IN}, then the output voltage will be less than the input voltage. If V_{Q1} will be greater than double of V_{IN}, then the output voltage will be greater than the input voltage.

B. DESIGN EQUATIONS

DUTY CYCLE
$$D = \frac{V_o}{V_o + V_{in}}$$
$$.....(6)$$
INDUCTANCE
$$L1 = L2 = \frac{DV_{in}}{f\ \Delta I_L}$$
$$.....(7)$$
CAPACITANCE
$$C1 = \frac{I_o\ D}{f\ \Delta V_{c1}}$$
$$.....(8)$$

$$Co = \frac{I_o\ D}{f * 0.5 * \Delta V_{c2}}$$
$$......(9)$$

C. DESIGN PARAMETERS

Input line voltage (V_s)   = 12V
Output voltage   = 48V
Output power   = 46W
Switching frequency   = 100 kHz
Duty cycle (d)   = 0.8

Line frequency (fs)   = 50Hz
Inductance (L1=L2)   = 110µH
Capacitance C_1 = 5µF
Capacitance C_2 = 300µF
Load resistance   = 50 Ω
Load inductance   = 110µF

IV. DESIGN OF SMC FOR SEPIC

The design of smc for sepic converter is obtained by deriving its mathematical model using state space averaging technique.

A. MATHEMATICAL MODELLING OF SEPIC

The advantage of mathematical modelling is to handle the system easily with several inputs and outputs. The system model includes the internal state variables and the output variable. The model directly provides a time-domain solution, which ultimately is the thing of interest. The effect of the initial conditions can be incorporated into the solution and the matrix modelling is very efficient from a computational standpoint for computer implementation.

The state variables of the sepic converter here considered are x1,x2,x3 and x4 which is inductor currents(iL1 and iL3) and capacitor voltages (Vc1 and Vc2) respectively.
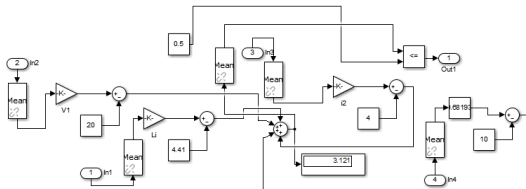
When the switch is turned on the state space equation can be obtained as

$$\dot{x}_1 = -\frac{v_{in}}{L_1}$$

$$\dot{x}_2 = \frac{v_{c_1}}{L_2}$$

$$\dot{x}_3 = -\frac{i_{L_2}}{C_1}$$

$$\dot{x}_4 = -\frac{v_{C_2}}{RC_2}$$

$$\frac{d}{dt}\begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{C_1} \\ v_{C_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{L_2} & 0 \\ 0 & \frac{-1}{C_1} & 0 & 0 \\ 0 & 0 & 0 & \frac{-1}{RC_2} \end{bmatrix} \begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{C_1} \\ v_{C_2} \end{bmatrix} + \begin{bmatrix} \frac{-1}{L_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} v_{in}$$

When the switch is turned off the state space equation can be obtained

$$\dot{x}_1 = -\frac{v_{C_1}}{L_1} - \frac{v_{C_2}}{L_1} + \frac{v_{in}}{L_1}$$

$$\dot{x}_2 = -\frac{v_{C_2}}{L_2}$$

$$\dot{x}_3 = \frac{i_{L_1}}{C_1}$$

$$\dot{x}_4 = \frac{i_{L_1}}{C_2} + \frac{i_{L_2}}{C_2} - \frac{v_{C_2}}{RC_2}$$

$$\frac{d}{dt}\begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{C_1} \\ v_{C_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \frac{-1}{L_1} & \frac{-1}{L_1} \\ 0 & 0 & 0 & \frac{-1}{L_2} \\ \frac{1}{C_1} & 0 & 0 & 0 \\ \frac{1}{C_2} & \frac{1}{C_2} & 0 & \frac{-1}{RC_2} \end{bmatrix} \begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{C_1} \\ v_{C_2} \end{bmatrix} + \begin{bmatrix} \frac{1}{L_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} v_{in}$$

The state space averaging technique,

$$\frac{d}{dt}\begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{C_1} \\ v_{C_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \frac{-1}{L_1} & \frac{-1}{L_1} \\ 0 & 0 & 0 & \frac{-1}{L_2} \\ \frac{1}{C_1} & 0 & 0 & 0 \\ \frac{1}{C_2} & \frac{1}{C_2} & 0 & \frac{-1}{RC_2} \end{bmatrix} \begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{C_1} \\ v_{C_2} \end{bmatrix} + \begin{bmatrix} \frac{v_{C_1} + v_{C_2} - 2v_{in}}{L_1} \\ \frac{v_{C_1} + v_{C_2}}{L_2} \\ \frac{-i_{L_1} - i_{L_2}}{C_1} \\ \frac{-i_{L_1} - i_{L_2}}{C_2} \end{bmatrix} \gamma + \begin{bmatrix} \frac{1}{L_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} v_{in}$$

$$\dot{X} = AX + B\gamma + C$$

The status of the switch is

$$\gamma = \begin{cases} 1 \to S \to ON \\ 0 \to S \to OFF \end{cases}$$

The system response is determined by the circuit parameters and coefficients k1,k2,k3 and k4. With a proper selection of these coefficients under any operating conditions, high control strength, performance and a quick sepic output response can be achieved.

V. RESULTS AND DISCUSSION

The main purpose of this section is to analyse the simulation studies of the sepic with four different control methods. The simulation are performed on the SEPIC 's circuit with the following conditions:

1. Without controller
2. With PI controller
3. With FLC
4. With Sliding mode controller

1. Without a feedback controller :

The SEPIC is simulated using a pulse generator which is connected to the MOSFET gate to give a pulse input with a frequency of 100KHz. It is found that in the absence of controller ,the voltage is not maintained with 48V. The duty ratio is 80% as calculated from the given specifications.
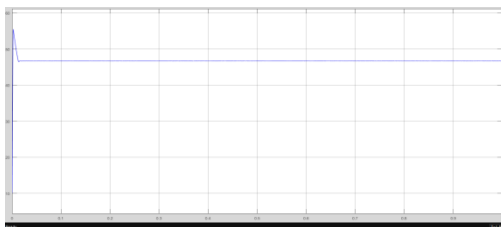
Figure 4. Open loop analysis of SEPIC converter

From figure 4 ,it is found that the voltage is regulated upto 45.31V in the absence of controller.

TABLE 1 : Performance analysis of SEPIC Converter with PI controller

% of load	Output voltage	Output power
100	48	46.08
75	48	38.4
50	48	23.54
25	48	17.01

2. With PI controller :

The Ziegler-nichols method is applied for determining the values of Kp and Ki. The values of Kp and Ki are 0.0001 and 0.1.
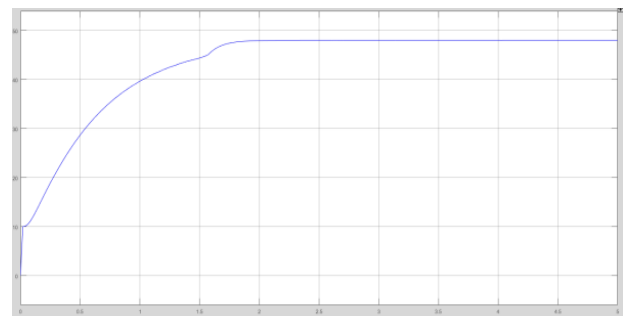
Figure 5. Closed loop analysis of SEPIC converter with PI controller.

3. With FLC :

One of the most commonly used soft computing technique is fuzzy logic based controller. The main purpose of FLC is to regulate the output voltage in the proposed system. Mamdani inference system is used here.
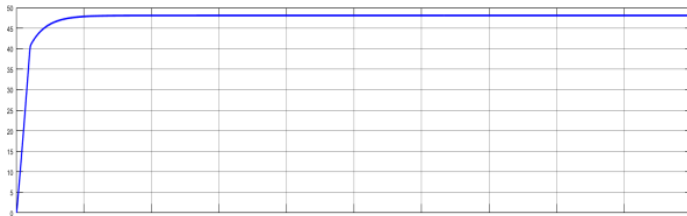
Figure 5. Closed loop analysis of SEPIC converter With Fuzzy logic controller.

% of load	Output voltage	Output power
100	48	45.1296
75	48	37.768
50	48	32.13
25	48	23.725

TABLE II.Performance analysis of sepic with FLC controller.

4.   With Sliding mode controller:

The parameters of the sliding mode controller which is calculated as per mathematical modelling of sepic are:

K1=K2=0.599,K3=0.378,K4=0.227.
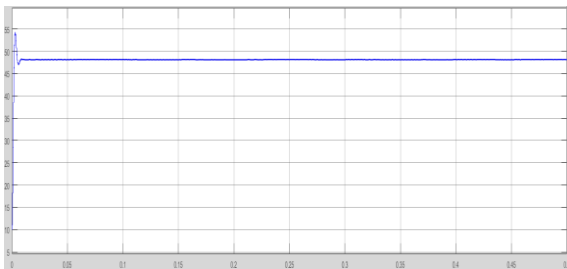
Figure 6. Closed loop analysis of SEPIC converter with SMC controller.

% of load	Output voltage	Output power
100	47.99	46.07
75	48.01	38.408
50	48.03	28.67
25	48.15	21.66

TABLE III. Performance analysis of sepic with sliding mode controller .

VI. CONCLUSION

The DC-DC SEPIC converter was simulated with variation in load. The voltage unbalance problem is improved by using proposed SEPIC converter. Mathematical modeling of the sepic converter is derived. The simulation was done for PI controller, fuzzy logic controller and sliding mode controller with variations in load. Fuzzy controller is easy to implement which provides crisp output and good robustness against load variation. Sliding mode controller for the sepic converter is proposed in order to ensure the stability under any operating conditions, better static and dynamic response under load variations. This system is suitable for real world applications lie power supplies for medical equipment, computer power supplies, uninterrupted power supplies, etc. The scope of the proposed system is suitable for real world commercial applications like power supplies for medical equipment, computer power supplies, uninterrupted power supplies, etc.

VII.   Acknowledgment

VIII.   References

[1]. SubhashChander, ,Pramod Agarwal and Indra Gupta, " Auto tuned discrete PID controller for DC-DC converter for fast transient response" 978-1-4244-7882-8/11/$26.00 ©2011/ IEEE.

[2]. Tadi G L Krishna Reddy,RamamurthyRaju.P, Revanthkumar.V, "Analysis of sepic for pv applications using PI and current mode controller", IJSRD-International Journal of Scientific and Research Development ,vol .1,issue 9,2013.

[3]. Amit Patel ,Kamal Singh ,Sandeep Kumar Singh, "Study and simulation analysis of single ended primary inductance conductor ", International Journal of Inventions in Reasearch, Engineering

Science and Technology (IJIREST),Vol.1,No.1,April 2014.

[4]. Abhinav Dogra, Kanchan Pal, "Designing and tuning of PI controller for flybackconverter",*International Journal of Engineering Trends and Technology (IJETT) – Volume 13 Number 3- Jul 2014*.

[5]. Eduardo I.ortiz-rivera, Jesus Gonzalez-Llorente, John E.salazar-Duque, "analysis and non-linear control of sepic dc-dc converter for photovoltaic systems", 978-1-4673-7653-5/15/$31.00 ©2015 /IEEE.

[6]. Venkatanarayanan Subramanian †and SaravananManimaran, " Implementation of a sliding mode controller for single ended primary inductor converter", Journal of Power Electronics, Vol. 15, No. 1, pp. 39-53, January 2015.

[7]. "Designing DC/DC converter based on sepictopology",by Jeff falin, power management corporation.

[8]. Designing a SEPIC converter, National Semiconductor Application Note 1484,

Wei Gu, Dongbing Zhang April 30, 2008

Design and implementation of control techniques for Zeta converter

Jajini. M[*1], Vairaprakash P[2]

*1PG Student, Department of Electrical and Electronics Engineering, Thiagarajar college of Engineering, Madurai, Tamil Nadu, India

2 Assistant Professor, Department of Electrical and Electronics Engineering, Thiagarajar college of Engineering, Madurai, Tamil Nadu,India

Abstract

The model proposed is Zeta voltage converter controller design and its control. The mathematical model of the zeta converter circuit operating in the continuous conduction mode is presented. PI controller, Fuzzy Logic controller and Sliding mode controller are the designed controllers. Analysis and comparison of simulation responses of open loop, close loop fuzzy logic controller and sliding mode controller are made with respect to periodic change in load. The results show that there is a significant performance and improvement in the proposed model.

Keywords : Zeta Converter; Voltage Regulation; State Space Model; Sliding Mode Control, Abbreviations and Acronyms : CCM-Continuous Conduction Mode;VRM-Voltage Regulator Module;SMC-Sliding Mode Controller;FLC-Fuzzy Logic Controller; D-Duty cycle; e-output error variable

I. Introduction

As the demand for high-performance, cost-effective systems continues to increase, power system design has become more challenging and complicated than ever before. The trend is DC-DC Converters with high efficiency and power saving capability.

India With the demand for portable electronics growing in India, the role played by DC/DC converters in reducing power consumption in hand-held devices is assuming greater importance. With more devices being designed so as to incorporate more features into them, portable system designers are looking at smaller and efficient DC/DC converters which perform efficiently whileincreasing battery life and system run-times. Improved efficiency, smaller form-factor, higher frequency switching and higher power densities are some of the trends that NXP is observing for DC-DC converters.

There are three main DC/DC converter technologies used with renowned applications. The first of these converters is the buck converter. Buck converters are step-down converters that output a voltage lower than the voltage that is input to the converter. The standard buck converter has an output that is equivalent to the input voltage multiplied by the duty cycle

$$V_{out} = D \times V_{in} \qquad \text{..... (1)}$$

Buck converters work for low voltage applications. The second commonly used converter is a boost converter. Boost converters are step-up converters that output a voltage higher than the voltage that is input to the converter. The standard boost converter has an output that is equivalent to the input voltage divided by the duty cycle.

$$V_{out} = \frac{V_{in}}{(1-D)} \qquad \text{.....(2)}$$

However, in many applications, a high boost ratio is required for the DC/DC converter to feed high voltage load or power grid. This cannot be satisfied by basic boost converters. Some limitation in conventional buck boost converter like inverted output, pulsating input current, high voltage stress make it unreliable for wide range of operation. So to get rid of this, ZETA converter is used.

The model implements Sinusoidal Pulse Width Modulation Technique to provide triggering pulses to the converter switch. With this, the output voltage will

be more with high modulation index. It also helps in reducing harmonics. Switch used is MOSFET, because it has higher speed of operation.

The controller action is provided by three different controllers. The PI controller acts as a comparative device that compares the input set predetermined value. If error is present, controller uses certain variables to correct them. The PI controller has negligible steady state error.

The controllers are fuzzy logic controller and sliding mode controller. Each control technique for the zeta converter is implemented using matlab simulation. The objective of the work is to stabilize and regulate the output voltage of converter under varying load conditions.

II. Zeta Converter

DC-DC conversion is the stem of Power Electronics.Many new converter topologies are still created every year. The Zeta converter is one among the new converter topology that provides a regulated output voltage from an input voltage varying above and below the output voltage.

Zeta converter is a fourth order DC-DC converter. The zeta converter shown in Figure 1 is capable of converting input voltage into a non inverted output voltage, having either a lower or higher value than input voltage. The inductor and the capacitors may also have large effects on the converter efficiency and ripple voltage. The energy transfer is controlled by high frequency switching device S (MOSFET).
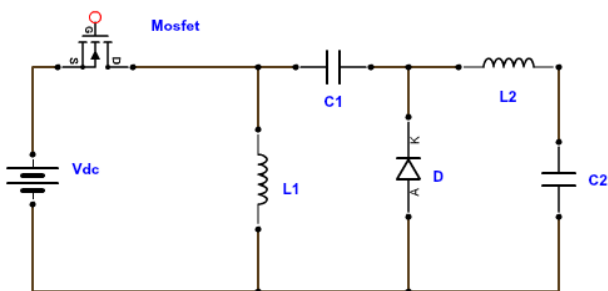
Figure 1 Zeta converter

A. Operating modes of zeta converter

Zeta converter is capable of operating in both continuous and discontinuous modes.The zeta converter consists of components like switch (S), inductors (L1 and L2), a diode, capacitors (C1 and C2), and a load (R).
MODE 1:

This mode is obtained, when the diode (D) is off and switch(S) is on. As shown in Figure 2 the current through the inductor L1 and L2 are drawn from the source Voltage V_s. The Inductor current i_{L1} and i_{L2} increase linearly. This mode of operation is called as charging mode.
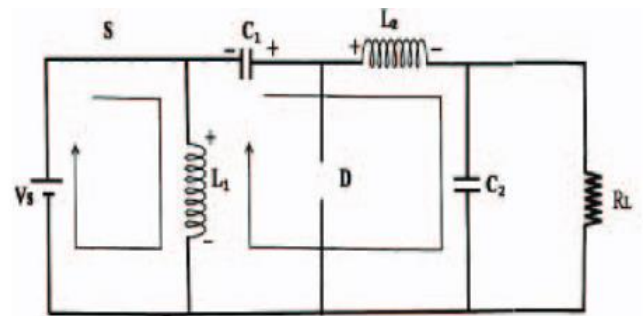
Figure 2 Zeta Converter in ON State

MODE 2:
This mode shown in Figure 3 is attained, when the diode (D) is in ON state and switch (S) is in off state. The energy stored in the inductors discharge and is transferred to the load. The current in the inductors decreases linearly. This mode of operation is called as charging mode.
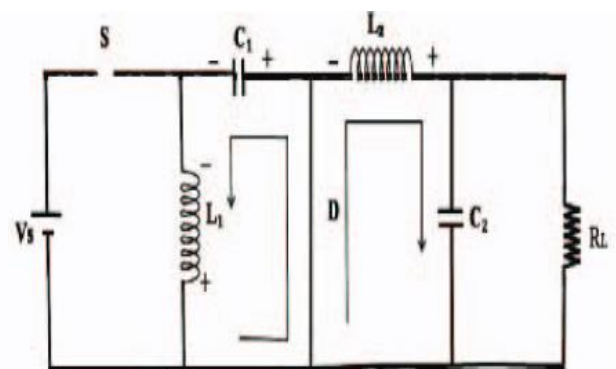
Figure 3 Zeta Converter in OFF State

Advantages of zeta converter:
- Unlike a classical buck–boost converter, the zeta converter has a continuous output current. The output side inductor makes the current continuous and ripple free.
- Although consisting of same number of components as Cuk converter, the zeta converter operates as non-inverting buck–boost converter. This property obviates a requirement of associated circuit for

negative voltage sensing, and hence reduces the complexity and probability of slowdown in system response.

B. Design equations of zeta converter

DUTY CYCLE

$$D = \frac{V_0}{V_0 + V_{in}} \quad \dots\dots(3)$$

INDUCTANCE

$$L1 = \frac{V_{in} D T_S}{\Delta I_{L1}} \quad \dots\dots(4)$$

$$L2 = \frac{V_0 (1-D) T_S}{2 I_0} \quad \dots\dots(5)$$

CAPACITANCE

$$C1 = \frac{V_0 D T_S}{R_L \Delta V_{C1}} \quad \dots\dots(6)$$

$$C2 = \frac{I_0}{2 w_L \Delta V_0} \quad \dots\dots(7)$$

III. Sliding Mode Controller Design

A. State space model of zeta converter

State-space averaging (SSA) is a reknownmethod formodeling switching converter.A state variable description of a system is written as follow

$$\dot{X} = Ax + BU$$

$$V_0 = Cx + DU \quad \dots\dots (5.1) \, \& \, (5.2)$$

Where A is n x n matrix, B is n x m matrix, C is m x n matrix and D is reserved to represent duty cycle ratio.

For a system that has a two switch topologies, the state equations can be describe as

When switch is closed

$$\dot{X} = A_1 x + B_1 U$$

$$V_0 = C_1 x + D_1 U$$

$$\dots\dots (8) \, \& (9)$$

When switch is open

$$\dot{X} = A_2 x + B_2 U$$

$$V_0 = C_2 x + D_2 U$$

$$\dots\dots(10) \& (11)$$

For switch closed at time DT and open for (1-D)T,the weighted average of the equations are

$$\dot{X} = [A_1 d + A_2(1 - d)]x + [B_1 d + B_2(1 - d)]U$$

$$V_0 = [C_1 d + C_2(1 - d)]x + [D_1 d + D_2(1 - d)]U$$

Let us assume the variables,

$$x_1 = i_{L1}$$

$$x_2 = i_{L2}$$

$$x_3 = V_{C1}$$

$$x_4 = V_{C2}$$

$$U = V_S$$

$$\dots\dots(12) \text{to} (18)$$

State space equation for on state is expressed as

$$
\begin{bmatrix} \dot{X_1} \\ \dot{X_2} \\ \dot{X_3} \\ \dot{X_4} \end{bmatrix} =
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{L_2} & \frac{-1}{L_2} \\
0 & \frac{-1}{C_1} & 0 & 0 \\
0 & \frac{-1}{C_2} & 0 & \frac{-1}{C_2 R_L}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} +
\begin{bmatrix} \frac{1}{L_1} \\ \frac{1}{L_2} \\ 0 \\ 0 \end{bmatrix} [U]
$$

$$\dots\dots(19)$$

State space equation for off state is expressed as

$$
\begin{bmatrix} \dot{X_1} \\ \dot{X_2} \\ \dot{X_3} \\ \dot{X_4} \end{bmatrix} =
\begin{bmatrix}
0 & 0 & \frac{-1}{L_1} & 0 \\
0 & 0 & 0 & \frac{-1}{L_2} \\
\frac{1}{C_1} & 0 & 0 & 0 \\
0 & \frac{1}{C_2} & 0 & \frac{-1}{C_2 R_L}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}
$$

$$\dots\dots(20)$$

The system state space equivalent equation is

$$
\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \\ \dot{x_4} \end{bmatrix} =
\begin{bmatrix}
0 & 0 & \frac{(1-d)}{L_1} & 0 \\
0 & 0 & \frac{d}{L_2} & \frac{-1}{L_2} \\
\frac{(1-d)}{C_1} & \frac{-d}{C_1} & 0 & 0 \\
0 & \frac{1}{C_2} & 0 & \frac{-1}{R_L C_2}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} +
\begin{bmatrix} \frac{d}{L_1} \\ \frac{d}{L_2} \\ 0 \\ 0 \end{bmatrix} [U]
$$

$$C = [0 \quad 0 \quad 0 \quad 1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$D = [0]$$

$$\dots\dots(21) \text{to} (23)$$

As= [0 0 (D-1)/L1 0;0 0 D/L2 -1/L2;(1 -D)/C1 -
D/C1 0 0;0 1/C2 0 -1/(R*C2)];
Bs= [D/L1; D/L2; 0; 0];
Cs= [0 0 0 1];
Ds= [0];
Vs= 12;
[Num, den]= ss2tf(As,Bs,Cs,Ds)
sys= tf(num,den)
step (sys*Vs)
grid

B. Determination of control parameters for SMC

The values of k is obtained by following equation

$$K_{eq} = B(BK^T)^{-1}K^TA \qquad \ldots\ldots(24)$$
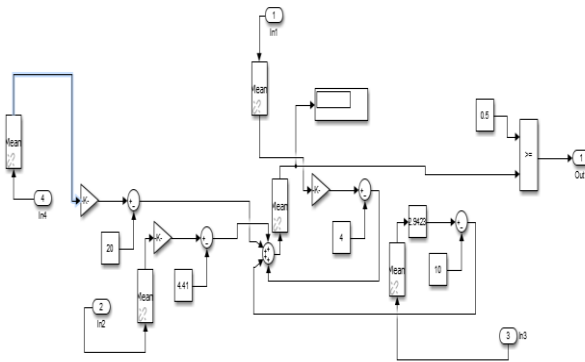
Figure 4 sliding mode control

IV. Simulated performance of proposed system

The dynamic performance of zeta converter is analysed
using following control techniques:
1. Without feedback controller
2. With PI controller
3. With fuzzy controller
4. With SMC controller

A. Zeta without feedback controller:

In open loop control of dc-dc zeta converter 20V input is
fed to which is to be boosted upto 48V as per
requirement and it is not obtained in open loop control.
A pulse generator is used to apply the gate pulse to the
power MOSFET. For open loop duty cycle chosen is
0.70

The desired output voltage 48V is not obtained in open
loop circuit. The obtained output voltage is 45.94V.

Figure 5. Open Loop Response

load %	R	OUTPUT VOLTAGE	INPUT CURRENT	OUTPUT CURRENT
100	23	48	11.7	2.0
75	30.6	48	8.4	1.5
50	46	48	5.3	1.0
25	92	48	3.2	0.5
10	230	48	1.0	0.2

TableI. performance analysis of converter for change in
load

B. Zeta with PI controller:

The proportional gain and integral gain values of PI
controller are obtained from Ziegler Nichols tuning
method. The proportional gain as 0.018 and integral gain
as 0.015 is taken to regulate the output voltage of
proposed converter.

In PI controller, the actual value of voltage and
thereference voltage are compared and the error value is
given to the PI controller. The output of PI controller is
then given to the relational operator where it is
compared with the repeating sequence. Here ramp signal
is considered.
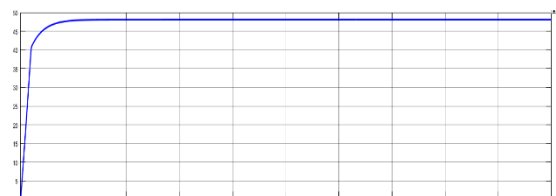
Figure 6. output voltage response with PI controller

load %	R	OUTPUT VOLTAGE	INPUT CURRENT	OUTPUT CURRENT
100	23	48	11.7	2.0
75	30.6	48	8.4	1.5
50	46	48	5.3	1.0
25	92	48	3.2	0.5
10	230	48	1.0	0.2

Table II. performance analysis of closed loop PI controller

C. Zeta with fuzzy logic controller:

The fuzzy logic controller is used to regulate the output voltage of proposed converter. The error signal and change in error signal are the inputs to the fuzzy logic controller.
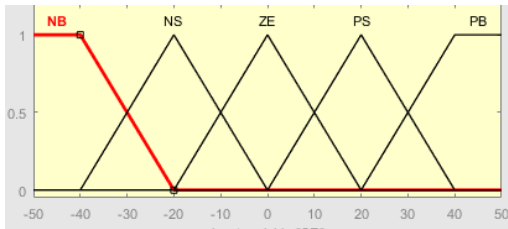
Figure 7. Input membership function (error signal)
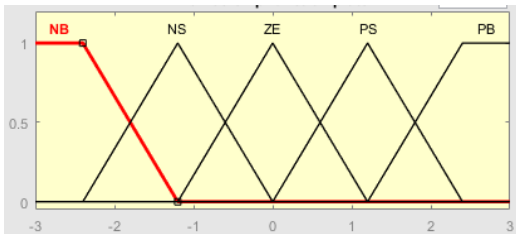
Figure 8. Input membership function (change in error signal)
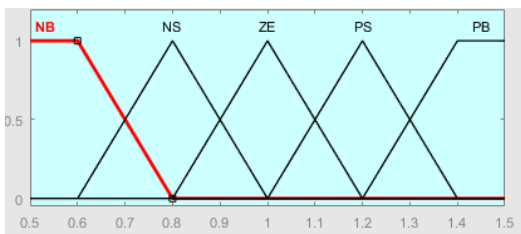
Figure 9. Output membership function
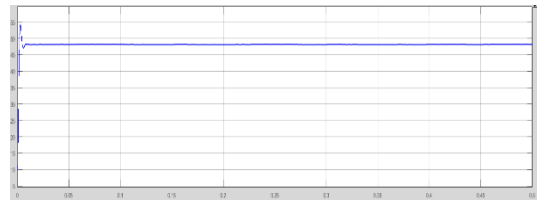
Figure 10. output voltage of Fuzzy Logic Controller

% of load	Output voltage	Output power
100	48.04	95.24
75	48.21	73.56
50	48.23	33.52
25	48.45	23.725

Table III. Performance analysis of closed loop flc controller

D. Zeta with SMC controller:

Here the current through inductors and voltage across capacitor are considered as the sliding surface. The gain values are calculated until the transient response is satisfactory.The measured capacitor voltages and inductor currents are made to compare with the reference values and the gain calculated are k1=k2=0.0599 , k3=0.0358 and k4=2.9423
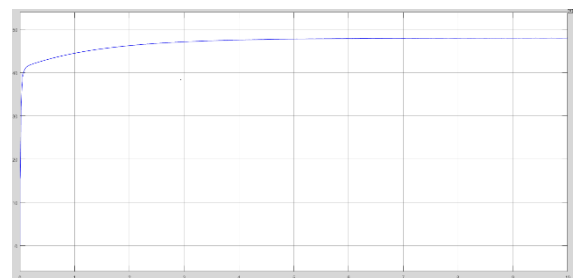
Figure 11. output voltage of SMC controller

% of load	Output voltage	Output power
100	48	97.17
75	48	73.508
50	48	65.38
25	48	24.68

Table IV. Performance analysis of closed loop smc controller

V. CONCLUSION

Zeta converter has been designed. Modeling and control of a Zeta converter operating in Continuous Conduction Mode (CCM) has been presented.The output voltage response as an open-loop system has been analyzed. The Zeta converter was simulated with variation in load. The simulation was done for DC-DC Zeta converter with the various controllers such as PI controller, Fuzzy logic controller and sliding mode controller to regulate the output voltage of the zeta converter circuit. The initiative is taken to develop sliding mode control technique for zeta converter.The computation time for the proposed system gains were very short. Hence the system gives better output and is efficient. The future scope of this work is usage of solar source employing effective MPPT algorithm

VI. Acknowledgment

VII. References

[1]. RajanKumar, Member, IEEE and Bhinsingh, fellow, IEEE "*bldc motor-driven solar pv array fed water pumping system employing zeta converter*" IEEE transactions VOL.52,NO.3,MAY/JUNE2016,PP.2315-2322.

[2]. Venkatanarayanan Subramanian †and SaravananManimaran, " Implementation of a sliding mode controller for single ended primary inductor converter", Journal of Power Electronics, Vol. 15, No. 1, pp. 39-53, January 2015.

[3]. SubhashChander, ,Pramod Agarwal and Indra Gupta, " Auto tuned discrete PID controller for DC-DC converter for fast transient response" 978-1-4244-7882-8/11/$26.00 ©2011/ IEEE.

[4]. J.falin, "Designing DC/DC converters based on zeta topology" Analog Applications Journal, Texas Instrumentsincorporated,pp.16-20,2010.

[5]. Ali H. Ahmad, Nashwan Saleh Sultan" *Design and Implementation of Controlled Zeta Converter PowerSupply*"American Journal of Electrical and Electronic Engineering, 2014, Vol. 2, No. 3, 121-128

[6]. AbhinavDogra, Kanchan Pal, "Designing and tuning of PI controller for flybackconverter",*International Journal of Engineering Trends and Technology (IJETT) – Volume 13 Number 3- Jul 2014*.

Intelligent Control of Photovoltaic System Using Fuzzy Logic for MPPT

V. M. Nandhana, Dr. S. Arockia Edwin Xavier
Department of EEE, Thiagarajar College of engineering, Madurai, Tamilnadu, India
nanswakeen3012@gmail.com

ABSTRACT

This paper suggests a method for the maximum power point tracking (MPPT) of a solar system under variable insulation conditions. A fuzzy Logic control (FLC) based MPPT technique is proposed to improve the efficiency of a standalone solar energy system. Fuzzy logic controller is applicable to a DC-DC converter device. The various design steps of fuzzy logic controller are presented along with its simulation.A fuzzy system for tracking the maximum power point of a PV system. A simulation model consists of PV panel, Buck converter, and fuzzy based MPPT developed. Maximum power can be tracking by using fuzzy logic controller has been simulated in MATLAB/Simulink.

Keywords: Fuzzy logic, photovoltaic (PV) Module, Maximum Power Point Tracking.

I. INTRODUCTION

The energy demand in the world has been increased due to growth in the industrial revolution. Fossil fuels available are gradually reducing due to the increasing population; therefore energy can be saved by usage of renewable energy sources. Renewable energy is a form of sustainable energy. Renewable energy sources available are wind energy, solar energy, etc.

By using photovoltaic system solar energy can be generated which is a most important renewable energy sources. There are several advantages in PV system such as fuel cost, clean, pollution free, little maintenance and no noise. The efficiency of the solar energy can be increased by using MPPT Technique.

In this technique the maximum power of the PV module can be extracted and it is delivered to the load. While extracting the maximum power of the PV module .the temperature and insulation factors are also taken into account. The objective of this is to work is to maintain the voltage of the PV system closer to the voltage at which maximum power is produced. Because of environmental pollution, global warming PV system is mostly preferred nowadays. But the efficiency of the PV should be increase. This can be done with the help of MPPT control. Voltage and current feedback back methods and P&O technique, incremental conductance method. The proposed work deals with the intelligent fuzzy based MPPT controller to improve the efficiency by tracking the maximum power from the PV system.

II. MODELING AND CHARACTERISTICS OF PV MODULE

Equivalent circuit model of photovoltaic is a major part to analysis the operation of device and the dynamic interactions. In this paper, PV module is described using various mathematical equations and Matlab/Simulink is used to simulate the PV module.

A. Equivalent circuit of PV module

The solar panel consists of many solar cells, by connecting those solar cells in series and parallel that can form a PV module. The single diode model consists of current source, diode and two resistors. Whereas two diode model also available but not consider here. The equivalent circuit of the PV cell is shown in figure.2.1
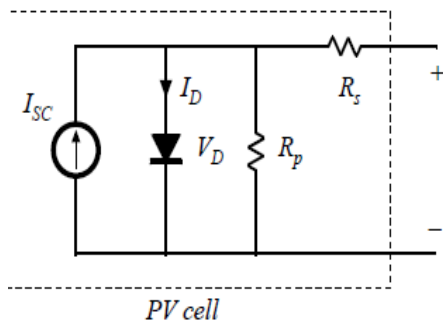
Fig.2.1 Equivalent circuit diagram of PV cell
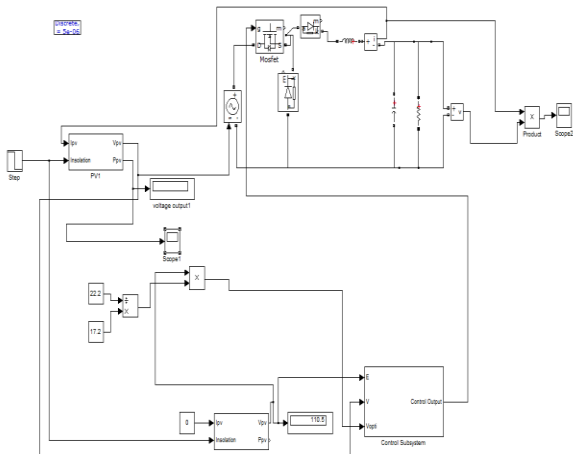
Fig.2.3 Simulink model of MPPT with FUZZY

B. Equations of PV module

From the theory of photovoltaic system, the following are the characteristic equations for a photovoltaic is given by

$$I_D = I_0(e^{\frac{V_D}{V_T}} - 1) \qquad (1)$$

$$I_{SC} - I_D - \frac{V_D}{R_P} - I_{PV} = 0 \qquad (2)$$

$$V_{PVcell} = V_D - R_S I_{PV} \qquad (3)$$

C. Matlab/Simulink modelling of PV module

The simulation model used for the implementation of the required PV module is as shown in figure.2.1.In this paper, I described current-input PV module, it has two inputs and two outputs. The .inputs are PV current Ipv [A] and insulation [W/m^2] and outputs are PV voltage [V] and PV output power Ppv [W]. This design is used for this method when the modules are connected in series, it shares the same current.Here temperature effect

is not designed. The data –sheet parameters are shown as below:

1. Short-circuit current Isc
2. Open-circuit voltage Voc
3. Rated current Ip at maximum power point(MPP)
4. Rated voltage VR at MPP
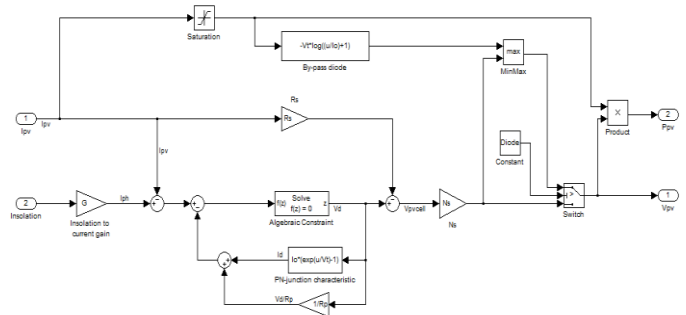
Fig.2.2 Matlab/Simulink model of PV module

III. NEED FOR MAXIMUM POWER POINT TRACKING

Due to changes in extrinsic and intrinsic factors in power voltage (P-V) and nonlinear current voltage (I-V) characteristics, we can achieve maximum efficiency at anytime in PV and also reduce the cost of energy. The PV system gives maximum power, when the maximum power point is unique point. There are many factor that can be considered, where the most important one is power point tracking in PV generation. The MPPT is nonlinear control problem. This is due to the nonlinearity present in the PV or parameter variations in PV.

To overcome this problem control methods like P&O, incremental conductance, neural and fuzzy are used. These control techniques requires high cost and these are complex. MPPT are simplicity and low cost, ease of implementation quick tracking under changing conditions, and small output power fluctuation.

IV. INTELLIGENCE MPPT TECHNIQUES

The photovoltaic system operation based on the load characteristics. When the direct connection between the source and load, the operating point is not optimal and output of PV module is soldem maximum. We use adaption device to overcome those problem that are MPPT controller with a DC-DC converter between source and load.

Whenever the insulation variation arises the MPPT controller track new modified maximum power point. The proposed Matlab/Simulink model of FUZZY based maximum power point tracking controller is depicted in Fig. 2.3

Many MPPT control technique are distinguish as:
1. Voltage feedback based method by comparing the reference voltage and PV operating voltage to generate PWM.
2. Current feedback based methods is done by computing short circuit of PV as a feedback to compute optimal current(maximum Power)
3. Power based method is used to track the maximum power point with the help of voltage and current of PV system.

A.FL-Based MPPT Technique:

Intelligent based MPPT shows better performance and lesser steady state error without any overshoot for fast changing temperature and irradiance. This proposed controller has two inputs and one output. One input is error (E) and another one is change in error (CE).

The following are the error (E) and change in error (CE) equations are shown as below:

$$e(k) = dP/dV(K) - dP/dV(K-1) \qquad (4)$$

$$C_e(K) = e(K) - e(K-1) \qquad (5)$$

Thus e (K) implies error at Kth instant. Fuzzy inference system used is Mamdani's and defuzzification is done by centre of gravity method, of this fuzzylogic-based MPPT as shown in Fig. 2.3. The rules are shown in below table.1

E / CE	NB	NS	Z	PS	PB
NB	NB	NB	NS	NS	Z
NS	NB	NS	NS	Z	PS
Z	NS	NS	Z	PS	PS
PS	NS	Z	PS	PS	PB
PB	Z	PS	PS	PB	PB

Table.1.Rules

V. RESULTS AND DISCUSSION

Simulation of PV module is done using fuzzy logic based MPPT controller is shown in section II .In this work DC-DC buck converter is connected between the load and PV for maximum power point tracking. Here fuzzy performance the control of duty cycle for buck converter in order to transfer maximum power to load and also comparison is done between PV with fuzzy and PV without controller.

The resulting outputs of voltage and current at insolation level of 900W/m2 are given in Figs. 5.1, 5.2 respectively. The proposed method results shows various from FUZZY based MPPT method shows several advantages when compare to other conventional methods of MPPT. FUZZY based MPPT method does not consume power in perturbation for maximum power point unlike the conventional P&O method. The major benefits of the presented MPPT method is that it can easily pick the rapidly changing environmental conditions without producing intrinsic steady state oscillations near the maximum power point.
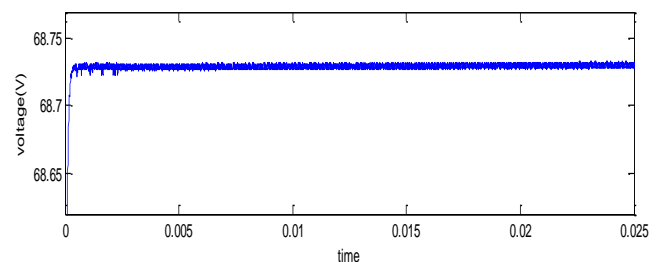
Fig.5.1 Output voltage vs time at 900W/m2 insolation
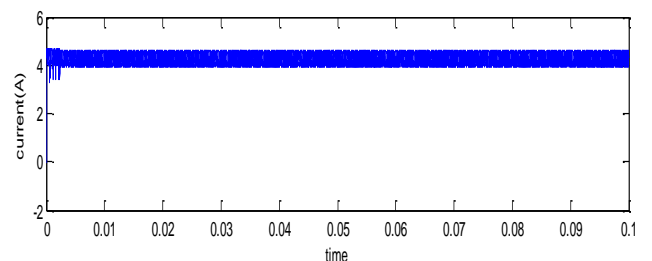
Fig.5.2 output current Vs time at 900w/m2 insolation

VI. CONCLUSION

From the above results it is inferred that MPP is tracked faster by using proposed MPPT method. Further it is observed that performance of the system is improved better by the implementation of FLC based MPPT

method. For non-linear system, this controller finds to be more effective than other controllers.

VII. Future work

Hardware implementation of the proposed FUZZY based MPPT controller by connecting Matlab with PV module and buck converter. An efficient MPPT controller of low cost and small size can be realized using a microcontroller and FUZZY based control scheme.

VIII. References

[1] Carlos A.P.Tavares1.Karla T. F. Leite2, Walter 1. Suemitsu3, Maria D. Bellar (2009)"performane evaluation of photovoltaic solar system with different MPPT methods," IEEE transactions pg978-1-4244-449.

[2] Huan-Liang Tsai, Ci-Siang Tu, and Yi-Jie Su, "Development of Generalized Photovoltaic Model Using MATLAB/SIMULINK", Proceedings of the World Congress on Engineering and Computer Science 2008 WCECS 2008, October 22 - 24, 2008, San Francisco, USA.

[3] G.J. Yu, Y.S. Jung, J.Y. Choi and G.S. Kim, 'A Novel Two-Mode MPPT Control Algorithm Based on Comparative Study of Existing Algorithms', Solar Energy, Vol. 76, N° 4, pp. 455 – 463, April 2004.

[4] H. Buhler, 'Réglage par LogiqueFloue', Presses PolytechniquesetUniversitairesRomandes, CH-1015, Lausanne, 1994.

[5] Jee-Hoon Jung, and S. Ahmed, "Model Construction of Single Crystalline Photovoltaic Panels for Real-time Simulation", IEEE Energy Conversion Congress & Expo, September 12-16, 2010, Atlanta, USA.

[6] K.M. Passino and S. Yurkovich, 'Fuzzy Control', Addison, Wesley, 1998._

[7] K. Punitha, D.Devaraj, S.sakthivel (2012),"development and analysis of adaptive fuzzy controllers for photovoltaic system under varying atmospheric and partial shading condition," science direct on applied soft computing vol.13, pp.4320-4332.

[8] L. J.L. Santos, F. Antunes, A. Chehab and C. Cruz, 'A Maximum Power Point Tracker for PV Systems Using a High Performance Boost Converter', Solar Energy, Vol. 80, N°7, pp. 772 - 778, 2006.

[9] M. Veerachary, T. Senjyu, and K. Uezato, 'Voltage-Based Maximum Power Point Tracking Control of PV Systems', IEEE Trans. Aerosp. Electron. Syst., Vol. 38, pp. 262 - 270, Jan. 2002.

[10] P. S. Revankar, W. Z. Gandhare and A. G. Thosar Government College of Engineering, Aurangabad, "Maximum Power Point Tracking for PV Systems Using MATLAB/SIMULINK", 2010 Second International Conference on Machine Learning and Computing.

[11] Ramos Hernanz, JA., Campayo Martín, J.J. Zamora Belver, I., LarrañagaLesaka, J., Zulueta Guerrero, E. Puelles Pérez, E., "Modeling of Photovoltaic Module", International Conference on Renewable Energies and Power Quality (ICREPQ'10) Granada (Spain), 23th to 25th March, 2010.

[12] R.Sridhar, Dr.Jeevananathan, N. ThamizhSelvan, Saikat Banerjee, "Modeling of PV Arrayand Performance Enhancement by MPPT Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 7– No.5, September 2010.

[13] Sandeep Kumar, Vijay Kumar Garg, "A Hybrid model of Solar-Wind Power GenerationSystem", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 2, Issue 8, August 2013, pp. 4107-4016.

[14] S.Chowdhury, S.P.Chowdhury, G.A.Taylor, and Y.H.Song, "Mathematical Modeling and Performance Evaluation of a Stand-Alone Polycrystalline PV Plant with MPPT Facility" IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, July 20-24, 2008, Pittsburg, USA.

[15] T. Takagi and M. Sugeno, 'Fuzzy Identification of Systems and Its Applications to Modeling and Control', IEEE Transactions on Systems Man and Cybernetics, Vol. 15, N°1, pp. 116 - 132, 1985.

[16] Vinodhini. R, Rajitha.G, Kiran Kumar B and prof.Arun Kumar .G,"dSPACE based 12/24v closed loop boost converter for low power application." VIT unersity 2014 International conference on computation of power, energy, Information and communication (ICCPEIC) 978-1-4799-382-11/14IEEE.

[17] X. Liu and L.A.C. Lopes, 'An Improved Perturbation and Observation Maximum Power Point Tracking Algorithm for PV Arrays', Power Electronics Specialists Conference, PESC' 04, 2004.

[18] Yi –Hua.Chun-Liu, Jia-Wei Huang, Jing-H siau Chen(2013)" neural network –based maximum power point tracking methods for photovoltaic systems operating under fast changing environments" Science Direct on solar energy 89 pg:42-53

[19] W.Xiao,W.G.Dunford,P.R.Palmer,A.capel, Application Of Centered differentiation And Steepest Descent To Maximum POWER POINT TRACKING,IEEE transactions on industrial electronics,vol.54,No.3 pp.2539-2549,may2007.

[20] C. Liu, B.Wu, and R. Cheung, "Advanced algorithm for MPPT controlof photovoltaic systems," in Proc. Canadian Solar Build. Conf., Montreal,QC, Canada, Aug. 20–24, 2004.

[21] N. Pongratananukul, "Analysis and Simulation Tools for Solar ArrayPower Systems," Ph.D. dissertation,Dept. Electrical and Computer Engineering,Univ. Central Florida, Orlando, FL, 2005.

[22] D. P. Holm and M. E. Ropp, "Comparative study of maximum powerpoint tracking algorithms," Progr. Photovolt.: Res. Applicat., vol. 11,no. 1, pp. 47–62, 2003.

[23] D. Shmilovitz, "On the control of photovoltaic maximum power point tracker via output parameters," Proc. Inst. Elect. Eng., vol. 12, no. 2, pp. 239–248, 2005.

Enhancing Social Video Services in Multi Cloud Using Distributed Graph

G. PrabuKanna, G. VijayaLalitha
Department of Information Technology, Kalasalingam University, Krishnankoil, Tamil Nadu, India

ABSTRACT

Cloud Computing is the emerging technology used to store and share data to various users. Online social network and videos which are generated by users are growing well in today's trend. Those social videos are hostedin cloud which is turning into a standard to give out the users, who generated and stored their contents in cloud. The crucial issue of cloud based video content sharing is that users found expansively, and they are not provided with good service provider. We proposed an algorithm in this paper, pixel permutation, to share social video to multi cloud providers and provide security to the video content as well as solution to the optimization problem. We exhibit the social video qualities and the tradeoff between them is uncovered for fulfilling the users and reducing the data transmission. Finally, the result is compared with the existing encryption algorithm and proved that the proposed algorithm is providing more security and the time required for processing is low when compared to existing algorithm. These are implemented using CloudSim and proved that proposed technique gives better result than the existing technique when compared to time and performance.
Keywords: Cloud Computing, Social Video, Cloud Service Provider, Pixel Permutation

I. INTRODUCTION

Now a day's the internet plays an important role in all places. Cloud computing is a term utilized for conveying the facilitated benefit over the web. Users are utilized to store their information in cloud. Cloud computing refers to the applications which are distributed as service over the Internet. Cloud is also defined as the datacenter with hardware and software. Cloud computing is used to use applications without installing it. With the help of Internet, users can access their datas and files anywhere and anytime.

Figure 1 represents the various challenges faced by cloud computing. Thus in this paper, need for security for the data stored in cloud and the allocation of the data to the cloud are presented in balanced manner and discussed below.

The information may be of text, image and video format. In future, video files will be progressively overwhelming today's traffic and all network traffic in 2018 will be video based. The encryption is the procedure used to scramble the information, which is shared to give

security. First, the key value is generated and shared between sender and receiver. The sender first scrambles the data, utilizing the key as of now produced, so that the other persons are not able to find the original text. The receiver on the opposite side can unscramble the substance just by giving the right key. This is mainly used for all multimedia content, for example, text, image and video. There are different techniques required to encrypt and decrypt the data. This paper presents an alternate cryptographic approach to offer security to the video files that are stored in cloud and transmitted between the sender and receiver.
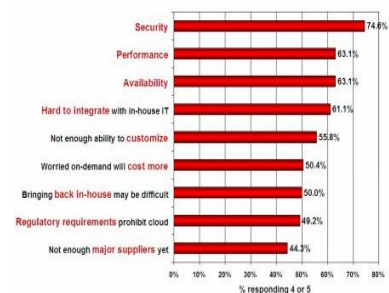
Figure 1. Cloud Computing Challenges

The various advantages of cloud computing are with the help of Internet the user can access, configure and manipulate applications. It is no necessary to install

specific software to access or manipulate applications. It offers on-demand services to the users. It also offers load balancing.

RESOURCE ALLOCATION TECHNIQUE

The main technique used in this paper is Resource allocation. It is the process of assigningexisting resources to the cloud through the Internet. It resolves the problem by letting the service providers to accomplish the resources for specific module.Resource allocation strategy is about integrating cloud provider actions for exploiting and allotting resources within the cloud environment. The below diagram represents the resource allocation strategy in cloud.
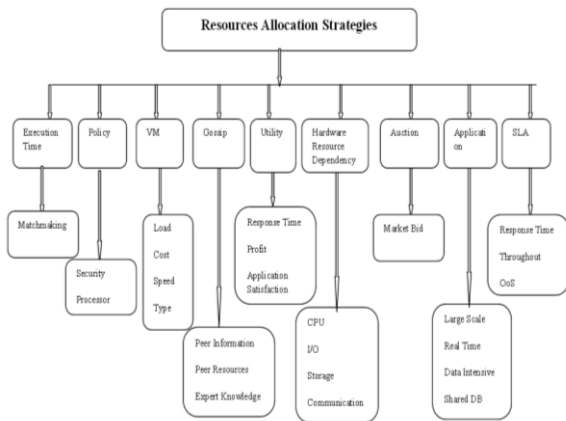
Figure 2. Resource Allocation Strategies

CLOUD COMPUTINGSECURITY ISSUES

Another process presented in this paper is cloud security issues. The issues confronted are:

1. Data issues

The first thing is the data in cloud can be retrieved anywhere and anytime by anyone. The service provider access and change the data using the data integrity methodin cloud. Then the second one is data stealing. Most of the users use other service providers because of its high cost. The third one is data loss. This is due to its financial or legal problem.

2. Privacy issues

The service provider makes sure that the datas are stored with high security and observing that who is maintaining and accessing the data.

3. Infected issues

The cloud service providers have all the rights to monitor and maintain the server. This will prevent malicious action by user from uploading wrong information on to the cloud, which may affect the services.

4. Security issues

The service provider should make sure that the server is provided with high security from all the external threats. Although, the cloud service provider should provide security for the client's information.
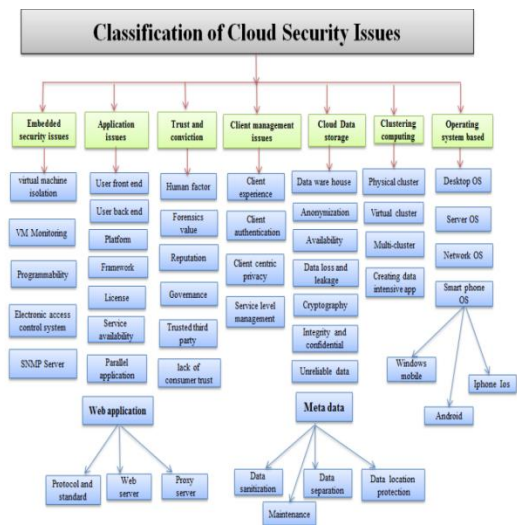
Figure 3.The classification of cloud security issues.

The below diagram represents the parameters that affect the cloud security issues.
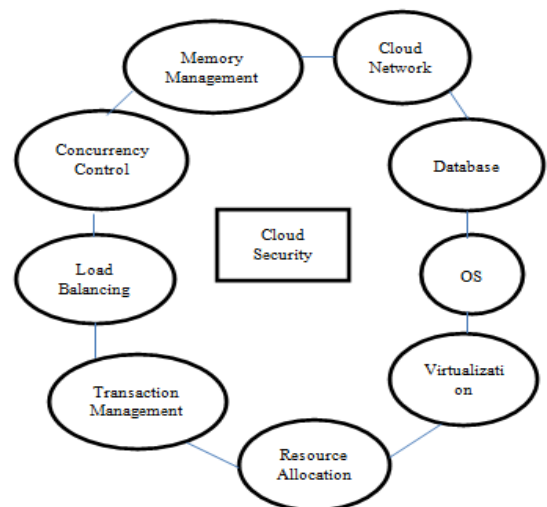
Figure 4. Parameters that affect cloud security

The security issues are briefly explained in [1], by Prince Jain. In this paper, the parameters which affect the cloud security and at the same way discover the security issues and problems faced by service providers are proposed. [2], by RojaRamani A, describe the data access control and data security in cloud and the architecture of the cloud,. Sharma et al., [3], proposed the description of various video encryption algorithms based on some parameters and difficult for particular algorithm to satisfy all parameter's performance. Omar et al., [4], proposed a description and comparison between encryption methods. Negi et al., [5], proposed various encryption algorithms for encrypting the video frames. Madhvi et al., [6], proposed a various encryption algorithms for video frames. The existing system proposes that the video frames are first encrypted using the standard tradition algorithm called Full Encryption Algorithm.The drawback of the full encryption is it is not appropriate for real time video frames because of its high computation and slow speed.

In this paper, we proposed a secure video sharing and storing process. The procedures involved are:

1. The size of the video frames is calculated using Random Walk Approach.
2. Then they are partitioned and stored to the specified cloud provider using Replacement Algorithm For Partitioning.
3. Then they are encrypted using Pixel Permutation Algorithm.

II. EXISTING WORK

A. FULL ENCRYPTION ALGORITHM

The existing system proposes that the video frames are first encrypted using the standard tradition algorithm called Full Encryption Algorithm. Hector et al., [9], proposed the networks performance and reduce the packet lossand SalahAly et al., [10], proposed various techniques for securing the multimedia contents.The process of encrypting each and every byte will be a slow and expensive one. The drawback of the full encryption is it is not appropriate for real time video frames because of its high computation and slow speed. While encrypting each and every frame it takes more time for processing. In existing work, the receiver is unable to have quality guarantee of an image. Thus the time

needed for full encryption technique is greater than the permutation technique. The users have to wait for long time to encrypt and decrypt the content. This leads to high cost.

B. PERFORMANCE AND RESULT

The results of the full encryption algorithm are shown below and it represents that the full encryption provides better result.

The original video frame before encryption process
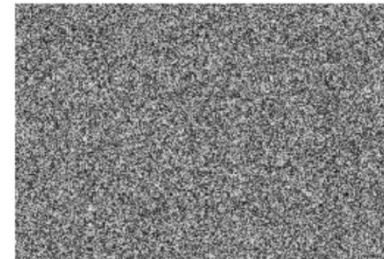
The video frame after encryption

The video frame after decryption

The drawback of the full encryption is it is not appropriate for real time video frames because of its high computation and slow speed. Thus the time needed for full encryption technique is greater than the permutation technique. The users have to wait for long time to encrypt and decrypt the content. This leads to high cost.

III. PROPOSED WORK

In this paper, first, the sizes of the video content are calculated and as per the frame size the video contents are hosted to the specified cloud provider using Random

Walk Algorithm. Then, the video contents are re-hosted to change the cloud providers using Partitioning Algorithm..Then the video files are encrypted and then stored in cloud usinga technique called, permutation algorithm. These are implemented using CloudSim and proved that this proposed technique gives better result than the existing technique when compared to time and performance. Thus the time needed for full encryption technique is greater than the permutation technique.

ALGORITHM STEPS

A. RANDOM WALK ALGORITHM

For re-hosting process the following algorithm is used. The random walk approach is one in which the highest node in the process is taken. Here node represents the video files and graph represents the deployment models. Then it is inserted as new node and the worst node is removed. This process is repeated when the value is less than its parameter value otherwise the process is stopped. The following equation (1) is used for the random walk approach.

$$p_{uv} = \begin{cases} \left(\frac{a}{m} + 1\right) \div (d_u + a), & \text{if } i \text{ and } j \text{ are linked} \\ \left(\frac{a}{m}\right) \div (d_u + a), & \text{if } i \text{ and } j \text{ are not linked} \end{cases} \quad \cdots$$

(1)

Table 1.Random Walk Algorithm

1. Assign n, a and x
2. Now random walk approach is executed according to the equation (1). If the first step is satisfied, uniform distribution is started.
3. The current node has the highest degree than the other node in the n list, insert new node and delete the worst node.
4. Return to 2 step if the random walk steps is less than x. Otherwise stop.

Here a and x are the parameters and the performance is based on these parameters.

We implemented the random walk approach for determining highest ranking node. This algorithm is not possible to apply ranking algorithm. In ranking algorithm, the calculation occurs at each and every graph of a global graph concurrently. The video files are stored in the cloud based on the sub graph placement algorithm. Then they assigned to store in to the various cloud providers. This process is based on the ranking algorithm. In which each files are assigned to various providers. Thus, it provides more security and does not allow other unauthorized persons to access the video files. Thus, the video files are re-hosted for changing the providers in cloud for clients and stored in cloud based on this algorithm. Finally they are encrypted using proposed pixel permutation algorithm.

B. REPLACEMENT ALGORITHM FOR PARTITIONING

Multi cloud means multiple clouds services in a single architecture. Multi cloud approaches can offer hardware, software and infrastructure redundancy to optimize fault tolerance. It can also navigate traffic from different clients through the internet. Instead of storing contents to a single cloud, multi cloud is used.When the whole content is stored in a single cloud provider, there is a possibility to hack the entire data by unauthorized persons. Thus the datas are partitioned and stored in multiple clouds to avoid unauthorized access to the content in cloud. The below diagram represents the multi cloud hosting. In single cloud, if whole server is cracked it is impossible to get back the content. But if we are using multi cloud there are various advantages. When one cloud provider is hacked or cracked it is possible to get the content back, because, if one server get cracked the remaining data may be retrieved from another server. So with the help of remaining content it is possible to get the original data.

In our proposed work, the multi cloud concept is used. First, the sizes of the video contents are calculated and then they are hosted to the specified cloud provider using replacement algorithm for partitioning. This is used to store the video contents to the cloud to partition the users to various cloud providers so that the videos are received with better quality.
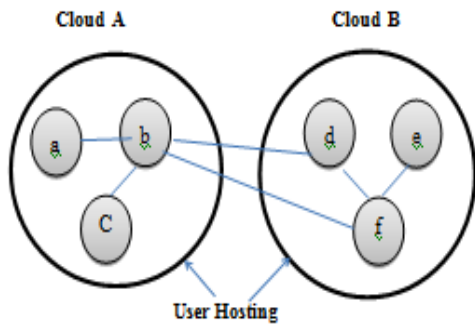
Figure 5. Multi-Cloud Hosting based on propagation.

Table 2.Replacement Algorithm for Partitioning

Here
a - Partitioned graph, p - Space in private cloud, c - Space in public cloud, t - Storage Multiplexer

```
result ←{}
for all ai in a do
        sortedP← sortDescending(P)
        if(sortedP.top().size()*t)>ai.size() hen
                sortedP.top().assign(ai)
                result.append(sortedP.top(),ai)
        else
                sortedC←sortDescending(C)
                if(sortedC.top().size()*t)>ai.size() then
                        sortedC.top().assign(ai)
                result.append(sortedC.top(),ai)
                else
                uNew←allocateNewPublicVM()
                        C.add(uNew)
                        goto 3
                endif
        endif
    end for
    return result
```

The sub graph replacement algorithm is shown in Table 2. Here the video files are sorted based on their sizes. In addition, the space in private cloud is sorted based on the available free space at that specified time. Then the leading space is allocated to the cloud, which is having the largest free space. After that, we sort them again based on the largest free space. We carry on this process until all of them are allocated to the cloud. In case, if the space is not enough to store the data, then we assign a new node in the public cloud and which is used based on the remaining space in the private cloud. The operation

of sub graph replacement algorithm is shown in Algorithm 1. This process is used to allocate the video files into the cloud.

C. PIXEL PERMUTATION TECHNIQUE

The Permutation Algorithm is one of the pseudo random generators in which a key is chosen randomly from a list of keys. The figure 6 represents the proposed permutation technique. The main purpose of permutation is to scramble only the particular video frames, not all the video frames are encrypted. SeshaPallaviIndrakanti and P.S.Avadhani in [7], and Hui et al., [8], proposed an image encryption for maintaining the quality of the imagebased on random permutation.

The pixel value of an image is chosen and permuted using the pseudo random index generator with the key already generated. These images are known as encrypted image. Now this encrypted image is transmitted to the receiver. The scrambled image is decrypted using pseudo random index generator and the same set of keys on the receiver side.
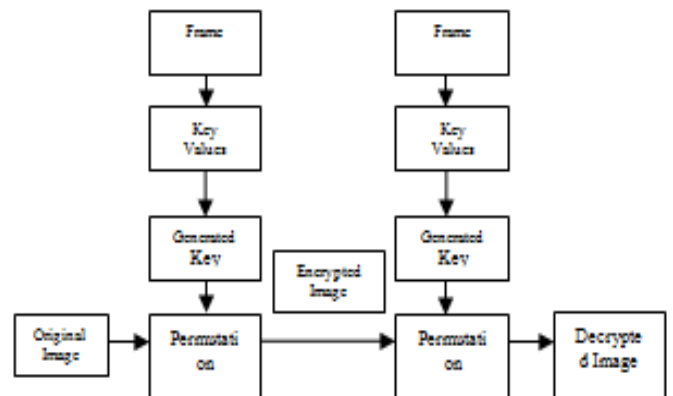
Figure 6. Permutation Technique

The following steps are used for implementing the pixel permutation algorithm.

Table 3. Proposed Pixel Permutation Algorithm
1. Choose the image and 8 bit key of an image is given as an input
2. Now the decimal pixel value of an image is converted into the binary value
3. The process is repeated for all the images' pixel value and based on the key value rearrange the bits.

4. Now the permuted value is converted to decimal value then form a matrix
5. Now the pixels are permuted based on the key and finally, they are arranged

The users have to choose the image and then encrypt. The key generated is used to encrypt and decrypt the frames. Then convert each frames decimal pixel value to the binary value and repeat the process for all the three planes. According to the key, rearrange the bits. Then convert the permuted value back to the decimal value then form a temporary matrix by transferring row of pixels, and permute pixels as per the key generated and then, partition the image into 8 blocks vertically or horizontally. Finally, rearrange the blocks based on the key generated.

IV. PERFORMANCE RESULT AND COMPARISON

This section determines the comparison of the proposed pixel permutation algorithm with the existing full encryption algorithm.
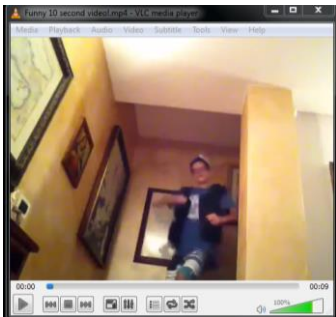
The original video frame before encryption process.
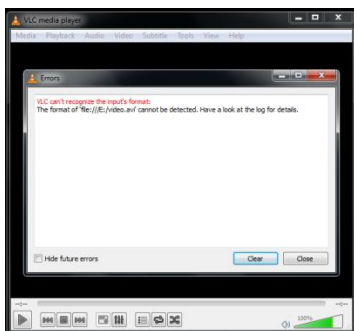
The video frame after encryption.
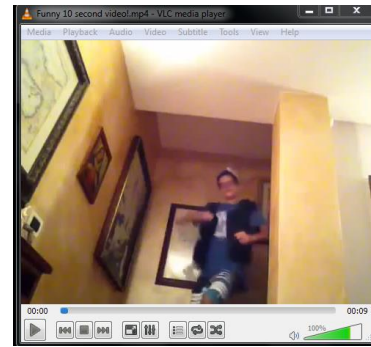
The video frame after decryption.

Here FEA – Full Encryption Algorithm and PPA – Pixel Permutation Algorithm.
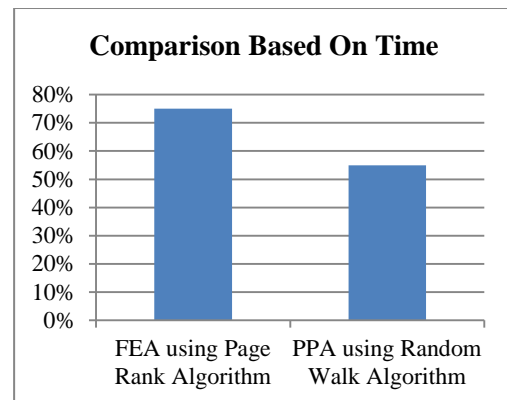
Comparison between Existing and Proposed Algorithm based on Time

The above comparison represents that the proposed pixel permutation algorithm is giving better result when compared to the existing full encryption algorithm based on time.
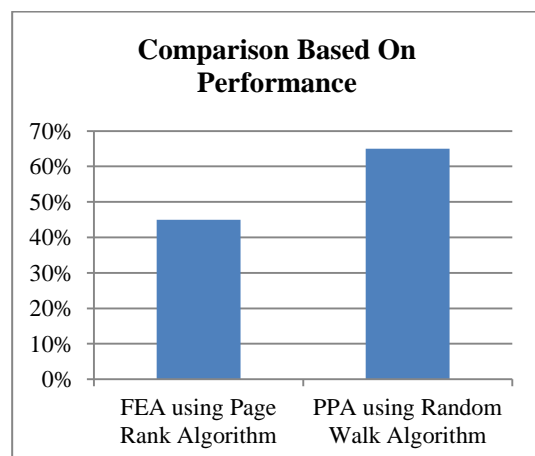
Comparison between Existing and Proposed Algorithm based on Performance

The above comparison represents that the proposed pixel permutation algorithm is better when compared to the existing full encryption algorithm based on performance.

V. CONCLUSION

In this paper, we studied hosting of the social video content with multiple providers in cloud. The problem in multi-cloud hosting is used to enhance users' satisfaction of provider preference in cloud and it is proved using Replacement Algorithm for Partitioning. Thus, the simple and effective methods are proposed in this paper for video content encryption using permutation techniques. The implementation is carried out using CloudSim. The result is compared with the existing encryption algorithm and proved that our algorithm shows better result based on both time and performance.

VI. REFERENCES

[1]. "Security Issues and their Solution in Cloud Computing", Prince Jain Malwa Polytechnic College Faridkot, Punjab-151203, India

[2]. "Encryption And Decryption Of A Cloud Computing File", RojaRamani A

[3]. "A Study Based on the Video Encryption Technique", Saurabh Sharma, Pushpendra Kumar Pateriya, Lakshmi *Department of Computer Science Engineering Lovely Professional University, Phagwara, India*

[4]. "An Overview of Video Encryption Techniques", M. Abomhara, Omar Zakaria, Othman O. Khalifa

[5]. "A Survey on Video Encryption Techniques", YogitaNegi*Asstt. Professor, BCIIT, Delhi* Haghighat, M., Zonouz, S., & Abdel-Mottaleb, M. (2015). CloudID: Trustworthy Cloud-based and Cross-Enterprise Biometric Identification. Expert Systems with Applications, 42(21), 7905–7916.

[6]. "A Survey of Video Encryption Methodologies", MadhviSoni , SapnaChaudhary, Dept. of Computer Science & Engineering, Shri Ram Group of Institutions, Jabalpur, India

[7]. "Permutation based Image Encryption Technique",SeshaPallaviIndrakantiP.S.Avadhani Department of CS and SE, Andhra University College of Engineering(A), Andhra University, Visakhapatnam

[8]. "Algorithm of Image Encryption based on Permutation Information Entropy", Guang-hui, Hu Kai, Yang He and E Xu, Beijing University of Aeronautics and Astronautics Beijing, China

[9]. "NAIVE – Network Aware Internet Video Encoding", Hector Briceno, Steven Gortler, Leonard McMillan, MIT.

[10]. "A Light – Weight Encrypting For Real Time Video Transmission", Salah Aly

[11]. "Known-Plaintext Attack Against a Permutation Based Video Encryption Algorithm", Adam J. Slagell slagell@ncsa.uiuc.edu January 16, 2004

[12]. "Digital Video Encryption Algorithms Based on Correlation-Preserving Permutations", Daniel Socek, Spyros Magliveras, DubravkoCulibrk, Oge Marques, HariKalva and BorkoFurht Florida Atlantic University, Boca Raton FL 33431

[13]. P. Mell and T. Grance, "The NIST definition of cloud computing," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. NIST SP 800-145, 2009.

[14]. W. Zhu, C. Luo, J.Wang, and S. Li, "Multimedia cloud computing," IEEE Signal Process. Mag., vol. 28, no. 3, pp. 59–69, May 2011.

[15]. D. D´ıaz-S´anchez, F. Almenarez, A. Mar´ın, D. Proserpio, and P.A. Cabarcos, "Media cloud: An open cloud computing middleware for content management," IEEE Trans. Consum. Electron., vol. 57, no. 2, pp. 970– 978, May 2011.

[16]. Srinavasin, Madhan (2012). "State-of-the-art cloud computing security taxonomies: a classification of security challenges in the present cloud computing environment". ACM ICACCI'.

[17]. "Identity Management in the Cloud". Information Week. 2013-10-25. Retrieved 2013-06-05

[18]. T. ElGamal, "A public key cryptosystem and a signature scheme basedon discrete logarithms," in Advances in Cryptology. New York, NY, USA:Springer-Verlag, 1985, pp. 10–18.

[19]. G. B. Algin and E. T. Tunali, "Scalable video encryption of h. 264 svccodec," J. Vis. Commun. Image Represent., vol. 22, no. 4, pp. 353–364,2011.

[20]. J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attributebasedencryption," in Proc. IEEE Symp. Security Privacy, 2007,pp. 321–334.

[21]. D. Naor, M. Naor, and J. Lotspiech, "Revocation and tracing schemes forstateless receivers," in Proc. CRYPTO, 2001, pp. 41–62.

[22]. L. Qiao and K. Nahrstedt, "A new algorithm for mpeg video encryption,"in Proc. 1st Int. Conf. Imag. Sci., Syst. Technol., 1997,pp. 21–29.

[23]. J. Ren, K. Zhang, and X. Shen, "Exploiting mobile crowdsourcing forpervasive cloud services:Challenges and solutions," IEEE Commun.Mag.,vol. 53, no. 3, pp. 98–105, Aug. 2015.

[24]. J. Ren, K. Zhang, and X. Shen, "Social aware crowdsourcing with reputationmanagement in mobile sensing," Comput. Commun., vol. 65, no. 1,pp. 55–65, Aug. 2015.

Noise Removal in EMG Signal Using Data Fusion Techniques

R.Mahalakshmi*[1], K.Rajeswari[2]

*[1] Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, TamilNadu, India
94maha3@gmail.com[1]
[2]Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, TamilNadu, India

ABSTRACT

One of the main challenges in processing the biomedical signals, such as ECG , EEG and EMG is noise removal as they are easily get affected by various noises arising from different environmental conditions. Filtering out the noise from EMG signal improves the accuracy and performance of signal processing systems. But in practice, it is very complicated to filter out noise from the desired EMG signals to obtain noise corrupted raw signal. This paper proposes a new data fusion techniques to reduce the effect of noise on electromyography signals, that are to be further processed to get the required information. The proposed method results in EMG signal enhancement when a corrupted emg signal with an additive white Gaussian noise is the only available information. The main idea is to utilize the kalman filter to remove the noise and enhance the performance of electromyography signals.

Keywords: EMG, Datafusion, Kalman Filter.

I. INTRODUCTION

The bioelectric potentials related with muscle action constitute the Electromyogram (EMG). These potentials possibly measured at the surface of the body near a muscle of concern or directly from the muscle by penetrating the skin with needle electrodes. while most EMG measurements are estimated to achieve an indication of the amount of action of a given muscle, or group of muscles, rather than that of an individual muscle fiber, the pattern is usually a summing up the individual action potentials from the fibers constituting the muscle or muscles being measured. EMG electrodes lift up potentials from all muscles within the range of the electrodes, and so potentials from nearby large muscles may hold up with attempts to measure the EMG from smaller muscles, although the electrodes are placed directly over the small muscles. The action potential of a given muscle (or nerve fiber) has a predetermined magnitude, apart from the concentration of the stimulus that generates the response. Therefore, in a muscle, the concentration with which the muscle works does not enhances the net height of the action potential pulse but does increase the rate with which each muscle fiber and the number of fibers that are activated at any given time.

The magnitude of the calculated EMG waveform is the immediate sum of all the action potentials generated at any given time. These action potentials happen in both positive and negative polarities at a given pair of electrodes, they sometimes add and at times cancel. Hence, the EMG waveform shows very much like a random-noise waveform, with the power of the signal a function of the amount of muscle activity and electrode position. EMG is used as a diagnostics tool for recognizing neuromuscular diseases. several methods have been studied in the past 3 decades to eradicate the noise from the emg signal [3], [5], [6].To eliminate the noise from emg signals, kalman filter is used. The Kalman Filter (KF) is a dominant tool in the analysis of the development of a dynamical model in time. The filter gives with a flexible manner to attain recursive evaluation of the parameters, which are most favorable in the mean square error sense. The properties of KF along with the simplicity of the derived equations make it precious in the investigation of signals.

The summary of the Kalman Filter, its properties and its applications is presented. More particularly, center on the application of Kalman Filter in the Electroencephalogram (EEG) processing, tackling

development of Kalman Filter such as the Kalman Smoother (KS) in the time varying autoregressive model. This model can be represented in a state – space form and the employment of KF offers with an evaluation of the AR constraints which can be used for the evaluation of the dynamic signal. The KF is an estimator with attractive properties like optimality in the Minimum Mean Square Error (MMSE). Later than its discovery, this estimator has been used in several fields of engineering. These are as follows: control theory, communication systems, speech processing, biomedical signal processing, etc. The KF is not only an estimator but also a learning method. The observations are helpful to learn the states of the model. The Kalman Filter is also a computational tool and may exist some problems as a result of the finite precision arithmetic of the computers.

The paper is organized as follows: Section II, presents the system model and signal generation. The two fusion techniques are applied to EMG signals in the existence of failures while a comparative performance under simulated noise conditions is given in Section III. Finally, in the results, Performance of these algorithms are discussed.

II. SYSTEM MODEL

The processing scheme of the overall emg signal is seen in the below figure,
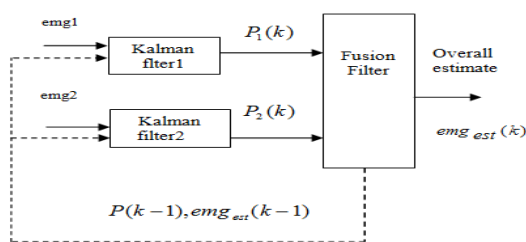
Fig.1. Outline of an overall fusion

In this two Emg signals were taken as the input signal to do the simulation. These two signals are obtained from adjacent muscles by two different electrodes.emg signals from the different electrodes are then fused using the variance weighted average to give the overall estimation of the signal.
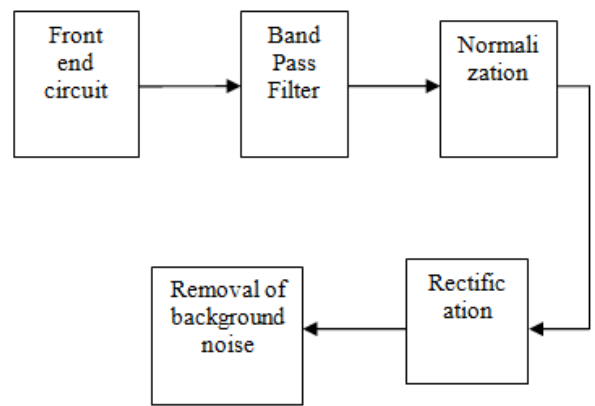
Above figure shows the signal generation of Emg. EMG is presented as a time sequence, which must be mapped to a smaller dimension vector by the computation of several features leading to a muscle force estimator and input to the classifier. A wide spectrum of features can be found in the literature, computed either in the time or frequency domain, or both, as can be seen in [13] and the references therein cited. Time domain features are widely used due to computational simplicity and real-time control possibilities. For choosing the most adequate, the statistical set proposed by [8] was evaluated in terms of computational cost and repeatability.

III. METHODOLOGY

There are Two proposed algorithms were used: Variance Weighted Average (VWA) and Kalman Filter (KF). while the EMG signal recorded during voluntary dynamic contractions can be considered as a zero-mean Gaussian process, modulated by muscle activity and corrupted by a Gaussian additive white noise [2], its instantaneous transformation of variance produces an indicator of muscle activity as well as the existence of fault-induced noise. For this basis, the variance was chosen as weighting function. In what follows, $emg_m(k)$ indicates the value of the EMG signal in channel i at the time of step k. With this value, the recursive computation of the instantaneous temporal mean (average) signal $\overline{emg}_m(k)(k)$ and the instantaneous variance $\sigma^2_{emg_m}(k)(k)$ was calculated for each sample time,

$$\overline{emg}_m(k) = \overline{emg}_m(k-1) + \frac{1}{k}(\overline{emg}_m(k) - \overline{emg}_m(k-1)) \quad (1)$$

$$\sigma^2_{emg_m}(k) = \sigma^2_{emg}(k-1) + \frac{1}{k}\left([emg_m(k) - \overline{emg}_m(k)]^2 - \sigma^2_{emg_m}(k)\right) \qquad (2)$$

A. Variance Weighted Average

A modified average was used in the first algorithm that is,

$$VWA(k) = w_0(k)emg_1(k) + w_1(k)emg_2(k) \qquad (3)$$

where signal weights are represented as $w_0(k)$ and $w_1(k)$ respectively, and represents the normalized coefficients, which are variable with the variance of the signals $\sigma^2_{emg_1}(k)(k)$ and $\sigma^2_{emg_2}(k)(k)$ in the time step k,

$$w_0(k) = \frac{\sigma^2_{emg_1}(k)}{\sigma^2_{emg_1}(k) + \sigma^2_{emg_2}(k)}, w_1(k) = \frac{\sigma^2_{emg_1}(k)}{\sigma^2_{emg_1}(k) + \sigma^2_{emg_2}(k)} \qquad (4)$$

where co-efficients $w_0(k)$ and $w_1(k)$ satisfy the ewfollowing conditions,

$$0 = w_0(k), w_1(k) = 1$$
$$w_0(k) + w_1(k) = 1$$

To understand this concept, we give an analytical example, i.e.,

If $\sigma^2_{emg_2}(k) \gg \sigma^2_{emg_1}(k)$ then VWA(k) $\cong emg_1$

If $\sigma^2_{emg_2}(k) \approx \sigma^2_{emg_1}(k)$ then VWA(k) $\cong \dfrac{emg_1 + emg_2}{2}$

A. Kalman filter

The Kalman filter processes data from many electrodes to provide a total state estimation in multi-sensor fusion. The Kalman filter can be used to combine or fuse information from different media or sensors and it is one of the Stochastic estimation tools. The Kalman Filter (KF) produces the overall estimate by reducing the variances [6]. Theoretically, in the system there is no performance loss, it delivers the same results as the centralized kalman filter, but the benefits of the KF are the modular concept that allows to add more sensors to the system, as needed, and an easier parallel implementation [2].

Figure 1 summarizes the concept of a KF, where the local filter outputs converge to the overall fusion filter via the respective variance and the estimated local outputs. In fact, as mentioned above, many local filters can be added as needed, and always the data are fused at the final filter.

As in the previous algorithm, the instantaneous variance is the decision parameter. Therefore, instantaneous mean and variance must be recursively computed with equations (1) and (2), thereafter; these values are inserted in the local filter (6). Finally, the vectors are fused in the overall filter (8), according to the procedure described by Soria et al. [4].

Based on the information available from $emg1(k)$ and $emg2(k)$, each kalman filter produce estimates (k) using the standard Kalman Filter equations. The Fusion Filter block fuses these estimates together to form the overall state estimate $emg_{est}(k)$.

$$P^{-1}_m(k) = P^{-1}_m(k-1) + (\sigma^2_m)^{-1} \qquad (5)$$

$$emg_{est_m}(k) = P_m(k)\left(P^{-1}_m(k-1)emg_{est_m}(k-1) + (\sigma^2_m)^{-1}emg_m(k)\right) \qquad (6$$

$$P^{-1}(k) = P^{-1}(k-1) + \sum_{m=1}^{n} P^{-1}_m(k) - P^{-1}_m(k-1) \qquad (7)$$

$$emg_{est}(k) = P(k)(P^{-1}(k-1)emg_{est}(k-1) + \sum_{m=1}^{n} P^{-1}_i(k)emg_{est_m}(k) -$$

$$P^{-1}_m(k-1)emg_{est_m}(K-1)) \qquad (8)$$

where m represents the local filter, n is the number of local filters, P_m stands for the local variance, (k) is the filtered (estimated) signal, P represents the global variance, and $emg_{est}(k)$ describes the global estimated vector.

IV. RESULTS

To achieve the aim of this paper the simulation is written in matlab. In this paper EMG signal is added with additive white Gaussian noise is chosen for simulation. Two emg signal from same muscle using different is taken for this experiment. Below figures shows that the

fusion techniques of kalman filter and variance weighted average produces a more precise estimation of emg signal from the input signal by filtering out additive white Gaussian noise.
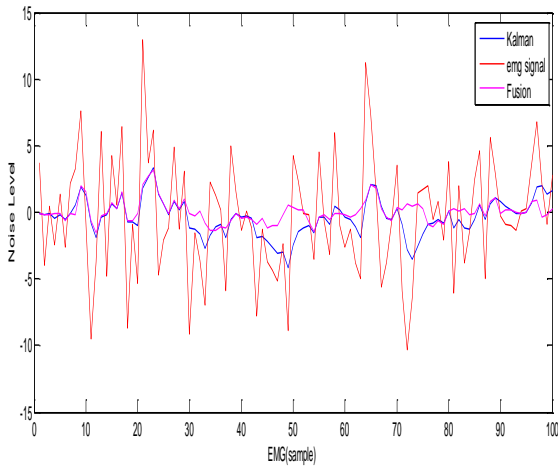
Fig. 3. First emg signal compared against fusion results

From figure (3), the first emg signal is compared with the kalman filter and fusion results, thus reduces the noise level. A graph will be drawn against the emg signal samples and the noise variance.
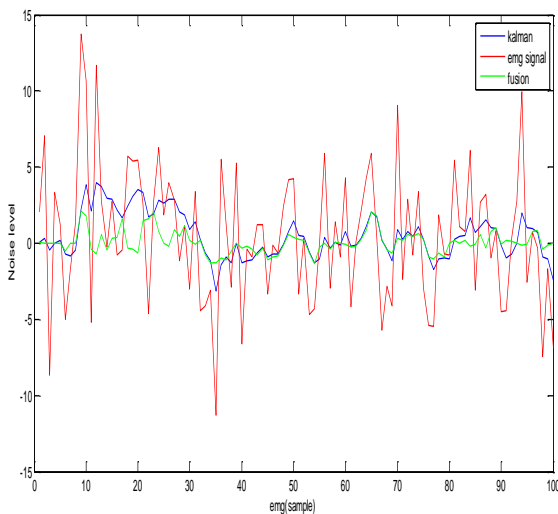
Fig. 4 . second emg signal with kalman filter and fusion.

Fig. 4. shows the second emg signal with the kalman filter and fusion results. Here, we have seen the fine difference between before and after the fusion. Kalman filter along with the variance weighted average method helps to remove the noise from the emg signal.
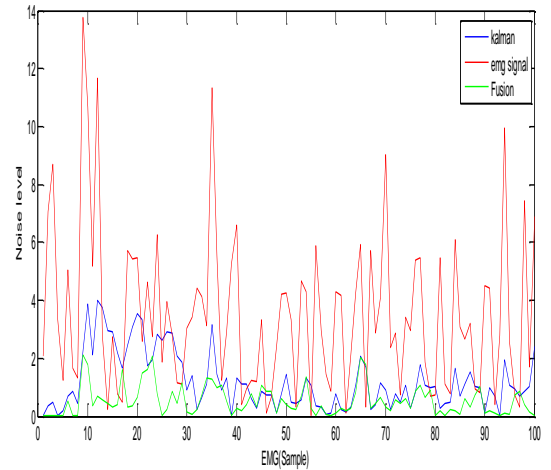
Fig. 5. Noise reduction using fusion at the positive co-

Ordinates

From the above figure, it is shown that absolute values of signals have been taken with respect to the positive co-ordinates.

Table 1: Maximum absolute error for fusion algorithms,

When the EMG signals were corrupted by Gaussian white noise.

Noise Level	White Gaussian Noise	
	Before Fusion	*After fusion*
0,1	0,972	0,174
0,2	0,710	0,242
0,3	1,094	0,286
0,4	0,693	0,416
0,5	0,692	0,403

When the benefits of each algorithm are investigated, as in any robustness scheme, two aspects must be taken into account: noise sensitivity and computational cost. On one hand, VWA appears as more efficient due to its better sensitivity to noisy signals, but the KF algorithm is suggested in fusion where the signals come from sensors whose nature is different, like electrodes for electromyography, piezoelectric contact for mechanomyography, accelerometers for acceleromyography (AMG), and condenser microphones for phonomyography (PMG). On the other hand, the computational cost for both algorithms is the same, therefore, this is not a decision factor, and the choice

would depend on the expected perturbations and the possibility of incorporating new sensors.

V. CONCLUSION

Two data fusion algorithms of EMG signals are proposed in this paper with the aim of improving performance of the emg signal. The main contribution of the work proposed here can be the improvement of emg signal under noisy conditions. The fact that two simple data fusion algorithms based on the variance analysis and without much computational cost were applied to EMG. The two algorithms used established an acceptable performance.

VI. REFERENCES

[1] Fukuda O, Tsuji T, Kaneko M, Otsuka A: A Human-Assisting Manipulator Teleoperated by EMG Signals and Arm Motions. IEEE Trans Robotics and Automation 2003, 19:210-222.

[2] Clancy E, Hogan N: Probability Density of the Surface Electromyogram and Its Relation to Amplitude Detectors. IEEE Trans Biomed Eng 1999, 46:730-739.

[3] Brown R, Hwang P: Introduction to Random Signals and Applied Kalman Filtering John Wiley & Sons; 1997.

[4] Soria C, Freire E, Carelli R: Stable AGV corridor navigation based on data and control signal fusion. Latin American Applied Research 2006, 36:71-78.

[5] F. Cattivelli and A. Sayed. Diffusion strategies for distributed kalman filtering and smoothing. IEEE Transactions on Automatic Control, 55(9):2069–2084, Sep. 2010.

[6] R. Olfati-Saber. Distributed Kalman filtering and sensor fusion in sensor networks.

[7] Oskoei M, Hu H: Myoelectric control systems. A survey. Biomedical Signal Processing and Control 2007, 2:275-294.

[8] Hudgins B, Parker P, Scott R: A new Strategy for multifunction myoelectric control. IEEE Trans Biomed Eng 1993, 40:82-94.

[9] Fleischer C: Controlling Exoskeletons with EMG signals and Biomechanical Body Model. In Ph.D. dissertation Berlin University, Germany; 2007.

Various Defect Detection Approaches in Fabric Images - A Review

M. Fathu Nisha*1, Dr. P. Vasuki2, Dr. S. Mohamed Mansoor Roomi3

*1ECE, K.L.N.College of Information Technology, Madurai, Tamil Nadu, India, mfnisha@rediffmail.com

2ECE, K.L.N.College of Information Technology, Madurai, Tamil Nadu, India, vasukip@klncit.edu.in

3ECE, Thiagarajar college of Engineering, Madurai, Tamil Nadu , India, smmroomi@tce.edu

ABSTRACT

Fabric defect detection is a necessary and essential step of quality control in the textile manufacturing industry. This paper has been reviewed the various fabric defect detection methods of statistical, spectral, model based and structural approaches. This paper has been presented the survey on types of defects, detection accuracy, performance metric and inference from recent publications. This paper shall benefit researchers and practitioners in image processing and computer vision fields in understanding the characteristics of the different defect detection approaches.The conclusion from this paper suggest that the pulse coupled neural network (PCNN) approach is better detection accuracy than the other methods and is suggested for further research.

Keywords: Quality control; fabric Inspection; Automated Defect Detection;

I. INTRODUCTION

Textile fibers can be made of natural element such as cotton or wool or a composite of different elements such as wool and nylon or polyester. Fabrics are textile materials which are made through weaving, knitting, braiding and bonding of fibers. Defect means a flaw on the fabric surface as a result of the manufacturing process. Fabric defect detection is a quality control process aimed at identifying and locating defects. Traditionally, human inspection, carried out in wooden board, is the only means to assure quality. It helps instant correction of small defects, but human error occurs due to fatigue, and fine defects are often undetected. Hence automated inspection becomes a natural way forward to improve fabric quality and reduce labor costs. Fabric inspection plays a significant role in quality control of textile materials.

Fabric defects are defined as distortions which occur on fabric pattern. Distortions cause to changing in Warp and Weft yarns, density of yarns or in spaces between Warp and Weft yarns. The automated fabric defect detection systems detect the extraordinary changes and decide the asset of defects and their places. Fabric manufacturing occurs in many stages, which start from manufacturing of yarns. In yarns manufacturing, fibers are processing for composing yarns. However, natural structure of the fibers or spinning operations may cause thinning, thickening, rupturing on fibers, these structural changes cause defects on fabric patterns.

After yarn manufacturing, weaving process composing fabrics with using yarns. The yarns are weaving by using looms. Sharpness parts of looms, abnormal motions of machines are also cause for defect reasons. All the defects that occur during these processes are named woven fabric defects. Fabrics are produced after passing many processing. Different machines and techniques are used during processing stages. So, fabrics are exposed to forces and stresses which cause defects. According to

their forms and directions, defects take different names has been extensively applied to various tasks and this method was reviewed in serial no of [1], [2], [4] in table 1.

B. Spectral approaches

The primary objectives of these approaches are firstly to extract texture primitives, and secondly
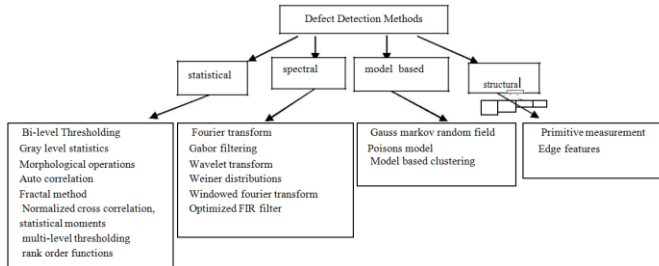
Fig. 1 Different defect detection methods

II. DIFFERENT APPROACHES OF FABRIC DEFECT DETECTION

A. Statistical approaches

Statistical method is a method of analyzing or representing statistical data. In fabric inspection, the main objective is to separate the image of the inspected fabric into the regions of distinct statistical behavior. Based on the number of pixels defining the local features, these approaches classified into higher order and lower order statistics. Higher order statistics (HOS) used the term skeweness and kurtosis whereas the low order statistics used the term mean and variance. Due to the higher powers, HOS are significantly less robust than lower order statistics. Clearly, the use of statistical approaches is well distinguished in the field of computer vision and to model or generalize the spatial placement rules. Spectral approaches require a high degree of periodicity thus, it is recommended to be applied only for computer vision of uniform textured materials like fabrics. For automated defect detection. Spectral approaches occupy a big part of the latest computer vision research work. This method was used for image quality enhancement, noise reduction, feature extraction and it was reviewed in serial no [3], [7], [8] in table 1.

B. Model-based approaches

In this approach, the texture is regarded as a complex pictorial pattern and can be defined by a deterministic model. Model-based texture analysis methods try to generate the texture. Here model the texture by determining the parameters of a pre-defined model.

Particularly, model-based approaches are suitable for fabric inspection when the statistical and spectral approaches have not yet shown their utility. The advantage of the modeling is that it can produce textures that can match the observed textures. Model based approaches are particularly suitable for fabric images with stochastic surface variations and this method was reviewed in serial no of [5], [9] in table 1.

C. Structural approaches

In structural approaches, texture is characterized by texture primitives or texture elements, and the spatial arrangement of these primitives. These primitives can be as simple as individual pixels, a region with uniform gray levels, or line segments. Consequently, the main objects of these approaches are firstly to extract texture primitives, and secondly to model or generalize the spatial placement rules. The Placement rules can be obtained through modeling geometric relationships between Primitives or learning statistical properties from texture primitives. However, these approaches were not successful on fabric defect detection, mainly due to the stochastic variations in the fabric structure and this method was reviewed in serial no [6], [10] in table 1.

TABLE 1
COMPARISON OF VARIOUS DEFECT DETECTION TECHNIQUES

S.No	Technique	Input image	Types of defects detected	Detection accuracy	Performance parameter	Inference
1	Fisher Criterion basedDeep Learning, Fisher Criterion Denoising Auto encoding(FCSDA) (Yundong le et al. 2016)	Plain, Twill fabric	Broken end, Hole, Netting multiple, Thick bar, Thin bar	98%	Learning rate, Scaling factor, Network depth, Number of hidden layer units	High detection accuracy was provided by this method even the negative samples are insufficient . Training process is time consuming.
2	Genetic Algorithm (Y.Kumbhar et al. 2016)	Textile color fabric	Threading defect, oil spot, color fading	90%	Homogeneity, contrast, Correlation, Hue, Saturation,	This method used to detect the defects for a sample of a very large number of high quality images.
3	Kernel Principal Component Analysis (KPCA) (Junfeng Jing et al. 2016)	Uniform & Structured fabric	Hole, Oil spot, Thread error, Objects on the surfaces	96%	Detection success rate specificity and sensitivity	High true detection rate and a low cost for online fabric inspection is achieved
4	Pulse Coupled Neural Network (PCNN) (Yundong Li et al. 2016)	Warp-Knitted fabric	broken ends, loom fly, thin bar	98.6%	Mean, standard deviation	This hybrid method is suitable to run on an embedded system because of the low computation. The system is applied to defect inspection for warp knitting machine.
5	GIMP Retinex filtering, (Amelio Carolina Sparavigana et al. 2016)	Textile woven fabric	Mispick	92%	Image size, scale division, dynamic slider	This method is used to help human vision for instance detecting tiny flaws of the fabric.
6	Feature extraction technique, graphbased	Color fabric	Needle cutting, oil	89%	Form factor, rectangularity	This method gives multipurpose without requiring any adjustment and finding the

	segmentation, (prasad dakhole et al. 2016)	image	stain		factor, location, orientation, intensity based features, moment based features	better accuracy and less time consumption in the industry.
7	2D Fast Fourier transforms (George bardi et al. 2016)	Woven fabric	Hole, excessive margin	78%	Yarn distance, energy distribution parameter, maximum energy parameter, displacement, speed	This method determines the yarn paths even in highly draped border regions of the surface. So the influence of different experimental settings on fiber orientation can be easily evaluated.
8	Small Scale Over Completed Dictionary Of Sparse Coding (SSOCD), Hardware Acceleration (Feng et al. 2016)	Twill fabric	Mispick, Broken pick, Buckling,Tom selvedge, Rough selvedge, Fuzzy ball, Bore, Yarn, Double flat, Yarn evenness, Pulled in selvedge, oil warp	93.7%	Reconstruction error, Sum of the absolute value of projection, largest absolute value of projection, distance between the image, average of trained samples, Size of image patches	Algorithm's serial time consumption is reduced. This system reached a state with a good parallel speedup ration and parallel efficiency. The algorithm run with high parallel efficiency and the detection speed meets the requirements of Industrial Inspection. the detection result will easily be affected by noise through, illumination changing, fabric grain fluctuating, etc.
9	Back propagation Algorithm (Akshay V.Nalawade et al. 2016)	Textile woven fabric	Oil spot, horizontal defect, vertical defect	88%	Entropy, homogeneity, contrast, wavelets, fractals	This method is reliable for fabric inspection system in textile Industries.
10	Artificial Neural Network back propagation pattern recoganization technique, series of filter operation (Gangandeep singh et al. 2016)	Woven fabric	3 types of defects	93.3%	5 features	The efficiency of this model is enough reliable to find the defects in fabrics. However this algorithm is not appropriate for all types of defects

III. CONCLUSION

Comparative studies in the literature that evaluate texture analysis methods in application to defect detection. In this paper FCSDA provide 98% accuracy and 2D fast Fourier transform provide 78% accuracy. It must be noted that different studies use different datasets and possible different parameter settings. The resolution of the images used for Various methods of defect detection of fabrics and quality and quantity metrics are reviewed in table1. There are several the detection process also important. Here the size of the image also affects the efficiency of the algorithm devised. Although a solid conclusion cannot be drawn to determine the best method for defect detection, statistical and filter based methods have been more popularly applied for fabric Inspection.

IV. REFERENCES

[1] Yundong Li, Weigang Zhao, Jiahao Pan, "Deformable Patterned Fabric Defect Detection with Fisher Criterion Based Deep Learning, IEEE Transactions on Automation Science and Engineering. (2016), 1545-5955,.

[2] Le Tong, W.K.Wong, C.K.Kwong. 2016 "Differential evolution based optimal Gabor Filter model for fabric Inspection", ELSEVIER, (2016), 0925-2312.

[3] Junfeng Jing, Xiaoting Fan, Penegfei Li, "Automated Fabric Defect Detection based on Multiple Gabor Filters and KPCA", International journal of Multimedia and Ubiquitous Engineering.(2016),Vol.11, No.6, pp.93-106

[4] Georg Bardi, Andreas Nocke, Chokri Cherif, Matthias Poch, Martin Schulze, Henning Heuer, Marko Schiller, Richard Kupke, Marcus Klein, "Automated detection of Yarn orientation in 3D-

draped carbon fiber fabrics and performs from eddy current data", Elsevier, 312-324, 2016.

[5] Akshay v.Nalawade, Pramod H.Narkhede, Swapnil B.Hande, Ganesh K.Sarade, Ranjeet S.Pisal, "Fabric Inspection System using Artificial Neural Networks", International Journal of Advanced Research in Electronics and Communication Engineering(IJARECE),Vol.5, Issue 1, 2016.

[6] Bhavini Patel, Hetal Bhaidasna, "Survey on Different Methods for Defect etection", International Research Journal of Engineering and Technology (IRJET), Vol.03, Issue:02, 2016.

[7] GangandeepSingh, Gurpadam Singh, Mandeep Kaur, "Fabric defect detection using series of Image Processing Algorithm & ANN Operation", Global Journal of Computers & Technology , Vol.4, No.2,ISSN:2394-501X, 2016.

[8] M.Aksha, V.Nalawade, Mr.Pramod H.Narkhede, Mr.Swapnil B.Hande, "Fabric Inspection System using Artificial Neural Networks", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Vol.5, Issue 1, 2016.

[9] Prasad Dakhole, Shruti kalode, Ankit Patel, "Fabric Fault Detection Using Image Processing Matlab" , International Journal for Emerging Trends in Engineering and Management Research(IJETEMR), -Volume II issue 1-21st January 2016.

[10] Dorian Schneider, Dorit Merhof, "Blind weave detection for woven ", Industrial and commercial application , springer, Pattern Anal Applic 18:725-737, 2015.

[11] I Gede Surya Rahayuda, "Texture Analysis on Image Motif of Endek Bali using K-Nearest Neighbor Classification Method", International Journal of Advanced Computer Science and Applications, Vol.6, No.9, 2015.

[12] Priyanka Vyas, Manish Kakhani, "Fabric fault processing using image processing techniques", International Journal of Multi disciplinary Research and Development , 2(2),29-31, 2015.

[13] Tiwari, Gaurav Sharma, "Automatic Fabric Fault Detection Using Morphological Operations on Bit Plane", International Journal of computer science and Network Security,Vol.15, No.10, 2015.

[14] Halil Ibrahim Celik, Mehmet Topalbekroglu, L.Canan Dulger, "Real Time Denim Fabric Inspection Using Image Analysis", Fibers and Textiles in Eastern Europe,23, 3(111):85-90, 2015.

[15] Lei zhang, Junfeng Jing, Hongwei Zhang, "Fabric defect Classification Based on LBP and GLCM", Journal of Fiber Bioengineering and Informatics 8:1, 81-89, 2015.

[16] Sudarshan Deshmukh.S, Raut.S, Birdar.M.S., "Defect Detection of regular patterned fabric by spectral estimation technique and rough set classifier", Proceedings of 45th IRF International conference , 978-93-85832-70-3, 2015.

[17] Tiangpeng Feng, Iian Zou, Jia Yan, Wenxuan Shi, Yifeng Liu,Clien Fan, Dexiang Deng, "Real time Fabric Defect Detection Using Accelerated Small Scale Over completed Dicitonary of Sparse Coding", International Journal of Advanced Robotic Systems, 13:1 Idoi: 10.5772/62058, 2015.

[18] Tianpeng Fen, Lian Zou, Jia Yan,Wenxuan shi, Yifeng Liu,Cien Fan, Dexiang Deng, "Real time Fabric defect detection using accelerated small scale over completed dictionary of sparse coding ", International Journal of Advanced Robotic Systems, 2015.

[19] Dandan ZHU, Ruru PAN, Wiedong GIO, Jie ZHANG, "Yarn Dyed fabric defect detection based on Autocorrelation function and GLCM", AUTEX Research Journal, DOI:1o.1515, 2015.

[20] Shweta Loonkar, Dr.dhirendra Mishra, "A Survey-Defect Detection and Classification for Fabric Texture Defects in Textile Industry", International Journal of Computer Science and Information Security(IJCSIS), Vol.13,No.5, 2015.

[21] Kumar Boyat, Brijendra Kumar Joshi, "A Review Paper: Noise Models in Digital Processing", Signal & Image Processing: An International Journal (SIPIJ), vol.6, No.2, 2015.

[22] Reza Ghazi saeidi, Amar Oukil, Gholam R.Amin, Sadigh Raissi, "Prioritization of textile fabric defects using ordered weighted averaging operator", SPRINGER, 76:745-752, 2015.

[23] Michael K.Ng, Henry Y.T.Ngan, Senior Member, IEEE, Xiaoming Yuan, Wenxing Zhang, "Patterned Fabric Inspection and Visualization by the method of Image decomposition", IEEE Transactons on Automation Science and Engineering, Vol.11, No.3,2014.

[24] H.Ibrahim Celik, L.Cannan Dulger Mehmet Topalbekiroglu "Fabric Defect Detection Using Linear Filtering and Morphological operations", Indian Journal of Fibre & Textile Research, Vol.39, pp.254-259, 2014.

Enhancement of Hybrid Power Scheme Based on Genetic Algorithm Using Three DC Source

*[1] U. Ramani, Department of Instrumentation and Control Engineering, Kalasalingam University, Virudhunagar, Tamailnadu, India
ramani.u6@gmail.com[1]

[2] Dr. B. Kannapiran, Department of Instrumentation and Control Engineering, Kalasaligam University, Virudhunagar, Tamilnadu. India
kannapiran79@ gmail.com[2]

ABSTRACT

Now a day's Solar Power is essential for all domestic purposes due to the demand of power consumption. But the single Solar system doesn't give the consistent yield due to sun power illumination. To conquer this issue, this project work presents enhancement of hybrid power scheme utilizes with PV/Fuel cell/Battery power to produce constant power using genetic algorithm (GA). The proposed hybrid scheme initially utilizes four module power controllers for independently expose to different duty cycles. The constant DC source is carried out by hybrid power scheme is to optimize the different duty cycle and to provide constant voltage through the load. This paper presents Genetic algorithm (GA) based model to enhance the duty cycle of the controller for a hybrid power control scheme. The simulation result was carried out by MATLAB/SIMULINK. From the model, the data required are generated and the performance is analyzed using a genetic algorithm. The GA based approach indicates better performance when compared with neural network controller.

Keywords: Hybrid DC source, Genetic algorithm, Duty cycle Enhancement.

I. INTRODUCTION

This document is a template. The standard fossil fuel vitality such as petroleum, natural gas, and coal are takes care of power demand in this day and age. The renewable vitality sources (solar, the wind, tidal, geothermal, etc.,) have more attention as an alternative vitality. In this paper, the photovoltaic (PV) vitality appears stopped to attract for electricity generation. Since, solar power grid generation relies on the sun illumination stage, encompassing temperature and eccentric darkness of PV [1, 2].

Fuel cell (FC) [2, 3, 4] exhibited to design the promising of extra power source that are advantages of neatness, more effectiveness, and prominent unwavering quality. For the most part, FC is perform extended turn on process and is carried out minimum time-consuming process of dynamic response to bungle power between the loads.

Batteries are for the most part taken a capacity component for charging and discharging power to improving the transitions of dynamic characteristics. Consequently, the three power scheme initialized by the DC source of PV/FC/Battery [5], which is known as hybrid power scheme. When evaluation individual sourced system as well as hybrid power scheme has to certainly offer by more quality output reliable and more efficiency. In this system consists of a storage feature, with utilize the power exist of bi-direction voltage flow. These three information control sources produce a decent capacity to supply voltage to the load alone.

Many hybrid power systems cannot able to maintain the constant output voltages for various power electronic converter application up to now. The ordinary techniques are used for individual power scheme provide smallest amount of voltage to overcome the problem evaluate with hybrid power scheme [6]. Nevertheless, the most traditional power scheme topology is very complex communicate system. Because, the performance time should be high together with large

amount of devices utilize the process provides high power losses. Here, quite a few stage of power conversion are utilizes the hybrid power scheme, which are changed by a three input converter. Thus, the result of process is consolidated through various power sources in individual structure. To defeat the problem suggested the boost converters having additional advantages of multi-objective optimization approach for hybrid power scheme by changing the duct cycle values randomized for developing the constant power deliver to the load. This paper deals with the multi-input dc-dc boost converter maintain the constant output voltage, to optimize the hybrid power system and fast output response using a genetic algorithm (GA).

Three input cuck regulated model fuzzy logic design [7] to interfacing the wind energy sources produce the electrical energy in to the system. This type of fuzzy the logic controller can be deliver yield streams relating to each power station and their reference voltage. DC-DC converter for battery [8] presents a controller is proficient to operating in bidirectional and to regulate the constant output voltage under different inputs values. The MPPT control algorithm present [9] that includes automatic determination and to permits the conventional converter of optimal operation control circuit under steady state environment with different duty cycles. DC-DC boost converters [10] present a design of integrated photovoltaic by applying back propagation artificial algorithm was produced boosted output voltage and sudden changes for solar illumination levels.

Modeling and design of regulated power supply approaches proposed [11] system conveying output source for converter topology with using a genetic algorithm (GA).This output impacts of parameters such like that increasing population size, iteration value, and varying the probability of mutation and probability of crossover also presented using MATLAB/SIMULINK.

Optimization of boost converter power loss was presented [12] a design of converter switching operation can be controlled, voltage swells by using an optimization technique. Utilizing these techniques easily obtained constant boosted output voltage by MATLAB. Hybrid power scheme optimization model were proposed for the design of [13] solar and wind energy system. This power scheme comprises of solar cells, wind turbines, and storage batteries are optimized using

objective function. This function optimized the total cost, initial cost, and yearly replacement cost can be maintained in the proposed converter.

Three input dc-dc converter was presented [14] fuel cells and battery packs are used to compensate for the immediate power changes in the microgrid that problem exhibits to the un even power loss of solar cells are required to the microgrid, and resultant of that power lose occurs sudden intrusion of the main grid. To overcome this losses using enhancement technique for three input dc source provide low level settling time and rise time in their grid connected system.

Optimization of hybrid energy system presented [15] an endeavour to convey the concept of the hybrid renewable energy system (HRES) and the condition of different values can be used to improve the simulink results. The enhancement techniques are used to reduce the swells and produce high output voltage for the microgrid application. Enhancement methods are utilized to reduce the investment cost and reliable resource.

II. DIFFERENT MODES OF OPERATION

The circuit comprise of a four MOSFET performs effectiveness of controlled with the independently depends on four duty ratios such as d_1 d_2 d3and d_4. Fig .1 performs with boost converter diodes D_1 and D_2 lead correlative way with comparing switches S_1 and S_2.

Referring this circuit when the switches S_3 and S_4 are get started with great operating mode and then possible of corresponding diodes D_3 and D_4 are biased followed by the complex criteria of battery voltage. These two switches turn-OFF state, the diodes $D_3$and D_4, ready to lead conduct the input currents iL_1, iL_2
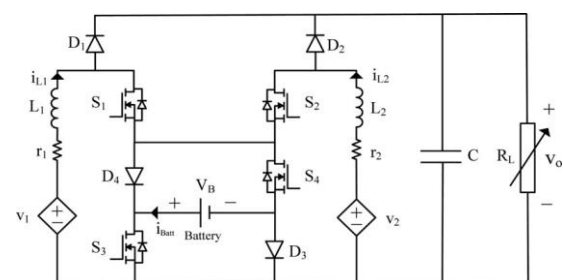
Figure 1. Circuit topology

In this circuit is hybrid power system applications and continuous current mode (CCM) is conceivable. It has the converter topology is seen from the solar or Fuel cell systems, an major problem is to reach an sufficient current swell and kept up the consistent yield voltage. The continuous current swells as well as input power are limited to make a correct power adjust along with the input powers simultaneously carried out the load. Consequently, the proposed hybrid scheme converter has been acquired the relentless state yield and quick yield reaction of the converter.

For the usually utilized condition occur of the storage element are considered by the DC modules with operation performs three modes, which are characterized to the proposed hybrid power scheme converters. These distinctive operation conditions are satisfied with d_3, d_4 <minimization (d_1,d_2) duty cycle ratios . While exceeding the process of duty ratio does not a failure in this converter. To defeat this issues always consider as the duty cycle ratio d_1 is smaller than duty cycle ratio d_2, then only switches are operated as autonomously.

A. *Mode1:*

In the mode 1, v_1 and v_2 charged power sources are providing through the load and without battery storage. That is fundamental modes are done in this mode to control the yield voltage. As unmistakably observed from the basic circuit topology, there are two approach to direct control of input power sources streams iL_1, iL_2. Without experiencing the battery;

❖ Current flow direction 1: S_4–D_3
❖ Current flow direction 2: S_3–D_4.

The working principal of mode 1 is feasible for applying the primary way. According this position switch S_4 is OFF and diode D_3 also OFF condition. The exchanging states are portrayed in Fig 2-4.

1) Condition state 1 ($0 < t < d_1T$)
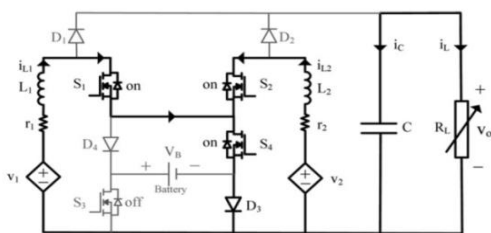
Figure 2. Condition State 1 ($0 < t < d_1T$)

In the main operation mode, at t = 0 the state of Switches S_1 and S_2 are turned ON until the inductors L_1 and L_2 are charged over the voltages v_1 and v_2 separately.

2) Condition state 2 ($d_1T < t < d_2T$)
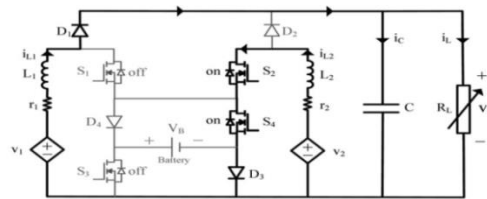
Figure 3. Condition State 2 ($d_1 T < t < d_2 T$)

In stage 2, the operation technique of t = d_1T are finished by Switch S_1 will to low state and this time switch S_2 is continuously ON position (if d_1 is smaller than d_2), the inductor L_1 is released to the charged voltage v_1-v_0, and output voltage is conveyed to the load. This circumstance inductor L_2 gets charged across the source v_2.

3) Condition state 3 ($d_2T < t < T$)
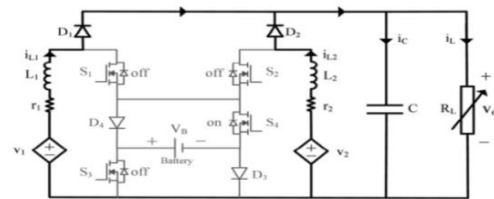
Figure 4. Condition State 3 ($d_2T < t < T$)

In stage 3, procedure of t = d_2T condition applicable of switch S_2 will be proceeds for quite little provided the voltage. While, the inductor L_2 is clearly decompose the voltage level beyond the v_2-v_0 voltage as similar to the inductor L_1. This operation is satisfied by the conventional theory as followed the voltage and current the associated by the condition are acquired with output response v_0 .

$$v_0 = \frac{v_1 - r_1 * i_{L1}}{1 - d_1} \qquad (1)$$

$$v_O = \frac{v_2 - r_2 * i_{L2}}{1 - d_2} \qquad (2)$$

$$C : (1 - d_1)Ti_{L1} + (1 - d_2)Ti_{L2} = T \frac{v_O}{R_L} \quad (3)$$

This operation mode is utilized to deal with the individual power sources deliver the input module goes on a kind of perspective power obtained by the duty

cycle values and the next power source is used to coordinate the output voltage.

B. Mode 2:

In the second mode, voltage source v_1 and v_2 along with the battery discharging state is incorporated for providing through the load. The basic converter circuit (Fig 1), the MOSFET switches S_3 and S_4 are turned ON. This procedure releasing is happened at whatever point the stream of current is between the i_{L1} and i_{L2}.

In any case, the discharging of the battery can just a single of the switches S_1 or S_2 is kept in ON position. This mode can be gotten by the resultant of maximum power. At this time duty cycle ratio diodes d_1 and d_2, the corresponding inductors current flow through the iL_1, iL_2.

Duty ratio d_4 as followed by the condition should be control the discharging power module of battery, which is concerning the certainties process continuously with occurs the mode switch S_4 operates with quite high condition. Then, it comes about passing through the currents iL_1, iL_2 provides for input belongs to the battery. As delineated in Fig 5-8, there are four attractive processes are operates by the DC converter module, as its individual exchanging state.

1) Condition state 1 $(0 < t < d_4T)$
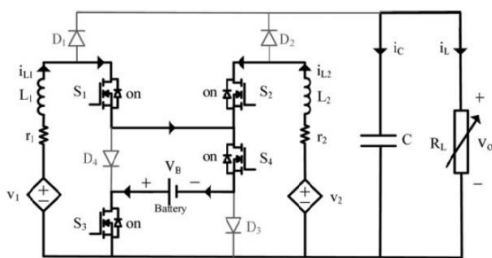
Figure 5. Condition State 1 $(0 < T < d_4T)$

In stage1 t= 0 switches S_1, S_2, and S_4 are absolutely greater than the other switches that instantaneously process of inductors L_1 and L_2 are existing improve the level across the voltages $v_1 + v_B$ and $v_2 + v_B$. These charged voltages are given to the load and create the steady output voltage for this switching state 1. Which state works according to the duty cycle ratio (supposition $d_1 < d_2$) used to turn ON the comparing switches.

2) Condition state 2 $(d_4T < t < d_1T)$
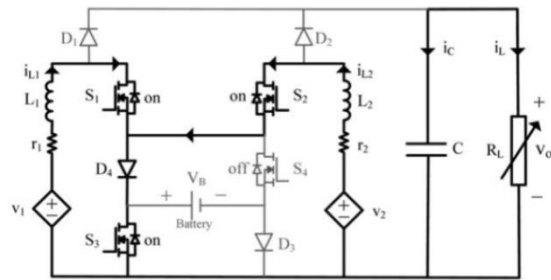
Figure 6. Condition State 2 $(d_4T < t < d_1T)$

In stage2 t = d_4T at this issues time switch S_4 is killed, while the switches S_1 and S_2 are quite with the higher than the others, which condition provides for inductors L_1 and L_2 are obtained with increase the voltage level across the v_1 and v_2 respectively.

3) Condition state 3 $(d_1T < t < d_2T)$
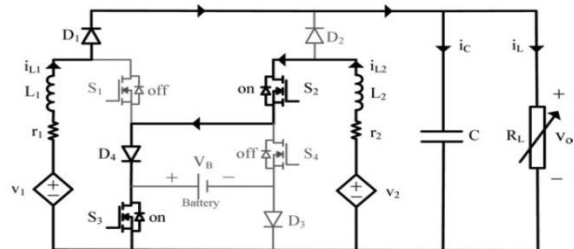
Figure 7. Condition State 3 $(d_1T < t < d_2T)$

In stage 3 t= d_1T at this condition switch, S_1 is execute with under low condition the meantime of inductor L_1 is simultaneously degrades the voltage beyond $v_1 - v_0$ at this stage L_2, is quite accused of voltages across v_2.

4) Condition state 4 $(d_2T < t < T)$
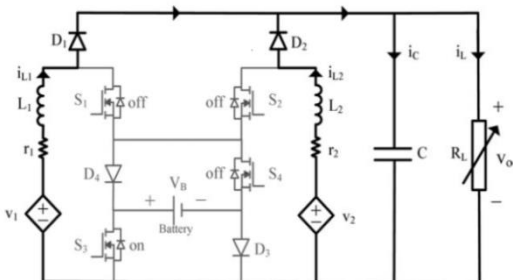
Figure 8. Condition State 4 $(d_2T < t < T)$

In stage 4 t = d_2T, this condition switch S_2 is likewise higher than the other switches, until the inductors L_1 and L_2 are simultaneously degrades the process over the voltage $v_1 - v_0$ and $v_2 - v_0$ individually. This operation is satisfied by the conventional theory as followed the

voltage and current the associated by the condition are acquired with output response v_0.

$$v_0 = \frac{v_1 - r_1 * i_{L1} + d_4 v_B}{1 - d_1} \quad (4)$$

$$v_o = \frac{v_2 - r_2 i_{L2} + d_4 v_B}{1 - d_2} \quad (5)$$

$$C : (1 - d_1)T i_{L1} + (1 - d_2)T i_{L2} = T \frac{v_o}{R_L} \quad (6)$$

In mode 2 is carried out by the v_1 and v_2 input power sources are referred to get the resultant of the duty ratios d_1, d_2 and the battery degrades as well as the power utilized balance output voltage by the limited duty cycle ratio d_4.

C. Mode 3:

In mode 3 provide for one or more different input source of v_1 and fuel cell v_2 are greatly extension of voltage as well as providing through the load and battery charging execution are incorporated into this mode.

As observed from the converter circuit, when the switches S_3 and S_4 are goes excessively turned OFF at these time switches S_1 and S_2 are quite higher than the others switches, then the inductor i_{L1} and i_{L2} provides the currents simultaneously followed by the access of diode D_4, battery and diode D_3. According to the condition battery is concurrently attended. The charging power is underneath the P_{max} esteem charging, to change the condition state of a one set of the switches S_3, S_4 earlier than the switches S1, S_2 goes to high state.

According to the condition state of this operation mode, when the switch S_3 is turned ON right now battery is holding off on charging until the switch goes too turned OFF state. These conditions state are illustrated in Fig 9-12.

1) Condition state 1 ($0 < t < d_3 T$)
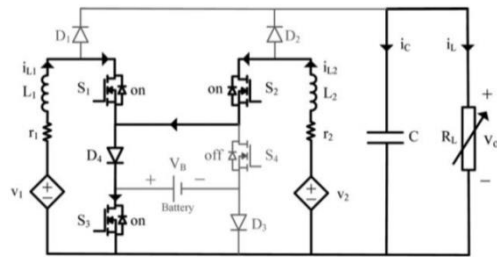
Figure 9. Condition State 1 ($0 < t < d_3 T$)

At this arrangement 1 t = 0 conditions that happens with the MOSFET S_1, S_2 and S_3 are greatly higher state compare to other states as shown fig.9.simultaneously the inductors L_1 and L_2 are accused by the v_1 and v_2, correspondingly.

2) Condition state 2 ($d_3 T < t < d_1 T$)
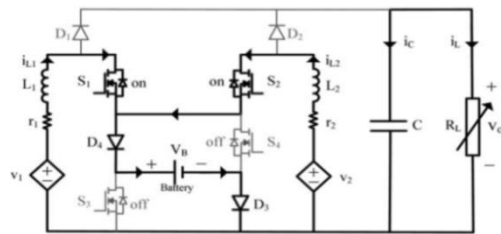
Figure 10. Condition State 2 ($d_3 T < t < d_1 T$)

In stage 2, t = $d_3 T$, this condition is fulfilled by low state of the switch S_3 and the MOSFET switches S_1, S_2 are at a standstill greater condition, and the inductors L_1 and L_2 to get charged through make use of the $v_1 - v_B$ and $v_2 - v_B$, separately.

3) Condition state 3 ($d_1 T < t < d_2 T$)
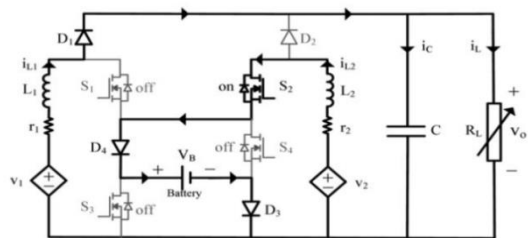
Figure 11. Condition State 3 ($d_1 T < t < d_2 T$)

In stage 3, t = $d_1 T$ condition takes after the switch S_1 is goes to low condition and the inductor L_1 is degrade across the voltage $v_1 - v_0$, at the similar time inductor L_2 is as yet charging across the voltage $v_2 - v_B$. . This operation is satisfied by the conventional theory as

followed the voltage and current the associated by the condition are acquired with output response v_0

$$v_0 = \frac{v_1 - r_1 i_{L1} - (d_1 - d_3)\, v_B}{1 - d_1} \quad (7)$$

$$v_o = \frac{v_2 - r_2 i_{L2} - (d_2 - d_3) v_B}{1 - d_2} \quad (8)$$

$$C : (1 - d_1)T i_{L1} + (1 - d_2)T i_{L2} = T \frac{v_o}{R_L} \quad (9)$$

4) Condition state 4 ($d_2 T < t < T$)
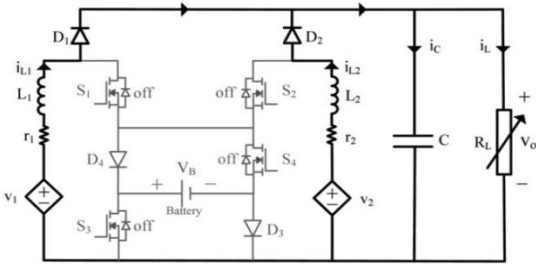
Figure 12. Condition State 4 ($d_2 T < t < T$)

In arrangement 4, t = $d_2 T$ condition is satisfied by the condition of low state switch S_2 and the inductor L_2, L_1 is released make use of the voltage $v_2 - v_0$.

III. PROPOSED METHODOLOGY

Genetic algorithms are the most specified analysis ideal approaches to solve a complex optimization problem. This approach can be extremely broad calculation and can function admirably in any pursuit space. The genetic algorithm is proficient to produce an exact the expert solution. This algorithm also produces several solutions to a given problem. GA is standout amongst the powerful methods with which to quick response to create an exact the expert solution for a specified problem. GA consists of the three frame work followed by the exact solution of essential population to arrange the many random values to evolve condition such as selection rules, crossover and mutation process as well as providing the better solution of the fitness function. This problem solve the required fitness function is minimizing the maximum rise time and settling time.
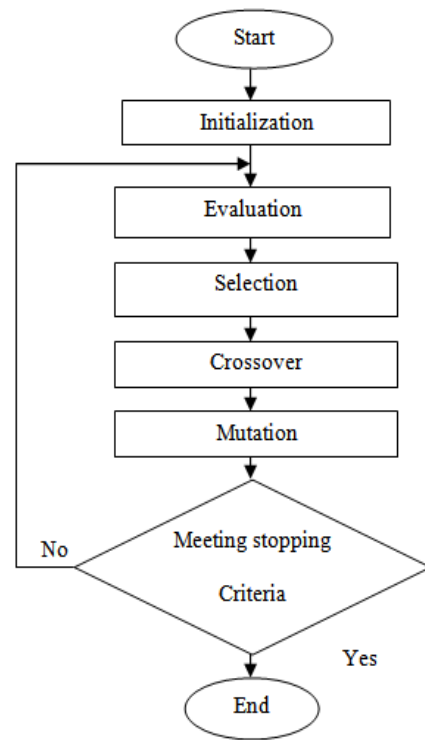
Figure 13. Flow chart of the genetic algorithm.

GA procedure based on the flowchart of fig 13. intially the GA select the parameters randomly for population size at initial stage and then produce the children from parents. This parent produces the children for next generation by using genetic algorithm flow chart.

IV. RESULTS AND DISCUSSION

In this paper, GA is used to optimize the duty cycle ratio in order to minimize the rise time and settling time. The boost converter duty cycle ratios are optimized by using MATLAB/Simulink. The optimized duty cycle values are obtained by using the objective function. Now the population size is 100, initial population size is 1 and iteration value is 10. After the iteration can able to satisfy the minimum rise and settling time to produce the constant output voltage for proposed converter. Generally, the dc-dc boost converter duty cycle objective function as follows:

$$D = \sum_{n=0}^{Ts} \frac{Von}{T} \quad (10)$$

Where *Ts* is the Rise time,

Von is the pulse on time and *T* is the total time period

$T = Ton+Toff$ .This duty cycle ratio values can be obtained by MATLAB/Simulation result. After the simulation duty cycle values are saved in workspace file. Whenever applying the gate pulse to the proposed converter and take saved duty cycle ratio values are consider. The best optimal solution is obtained by changing the population size. Optimization results after 10 iteration as follows as:

Parameter	Values
Number of population size	100
Crossover Probability	1.0
Mutation probability	0.00001
Fitness value	0.006

Table 1. Parameters used in GA

This table 1.shows the parameter used in GA in order to optimize the duty cycle ratio. By selecting the population size as 100, the crossover probability as 1.0, mutation probability as 0.00001and the fitness value as 0.006. The simulation block of GA used for duty cycle optimization using MATLAB is shown in fig 14.
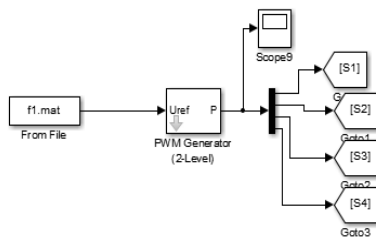
Figure 14. Duty cycle optimization using MATLAB.

Fig14.optimized values are given to the pulse generator and produce the duty cycle waveform for corresponding converter circuit in order to optimize duty cycle ratio.
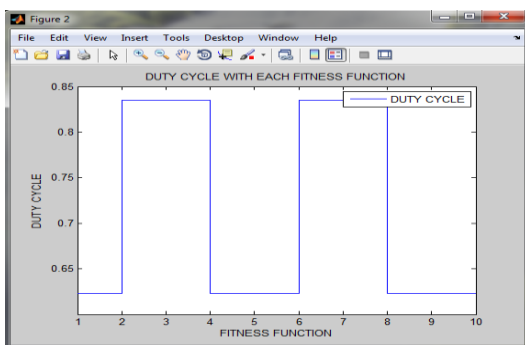
Figure 15. Duty cycle optimization Simulation

The duty cycle is optimize in the range of 0.85 is shown in Fig 15 with the given fitness values of 0.006 the duty cycle is optimized and by changing the fitness values of duty cycle can also be varied. This boosted output voltage waveform produces only after the 10th iteration and corresponding duty cycle ratio is used in this converter with the given values, the duty cycle is optimize as 0.85 and the corresponding boost output voltage using GA is shown in Fig 16.
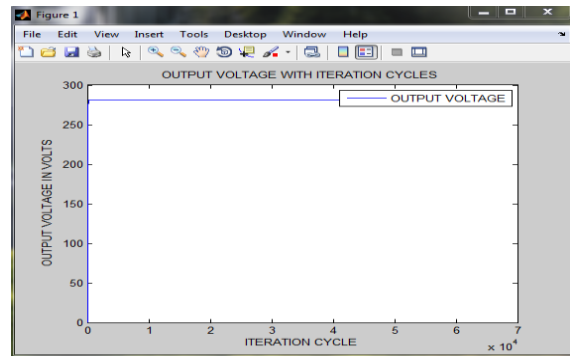
Figure 16. Boost output voltage using GA

V. COMPARISION OF ANN AND GA SIMULATION RESULTS

The GA model provides for duty cycle optimization of DC-DC boost converter from hybrid power scheme. The rise time required is 0.0005sec and settling time 0.009sec. For performance comparison ANN model was developed the rise time 0.005sec and its settling time 0.01sec. On comparing the performance of GA model with ANN model, the GA model shows better performance when compared with ANN model. Fig 17 is obtained the constant voltage level proceeding with proposed method under different operation mode for a hybrid power system using GA and also obtained better performance of settling time (Ts) and rise time (Tr), which this condition to maintain the continuously constant power delivered to the load.
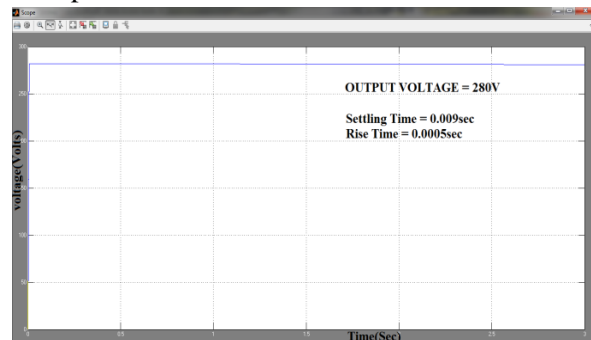
Figure 17. Boost output voltages using MATLAB/Simulink

The same problems were solved by the artificial neural network controller to obtained less settling, rise time and could not able to control the hybrid power scheme. But compare the genetic algorithm (GA) results is able to control the system and also obtained better optimization performance are given table 2.

Parameters	ANN	GA
Output Voltage	273.5V	280V
Tr	0.005sec	0.0005 sec
Ts	0.01 sec	0.009 sec

Table 2. Comparison between ANN and GA

VI.CONCLUSION

A hybrid power scheme of dc–dc boost converter structure was designed and better performance was obtained using GA. This hybrid power scheme can be apply for a solar (PV), FC, battery storage system and enhancement duty cycle ratios values are applied to the proposed converter circuit. This converter produces the yield output voltage source is obtained by using the Genetic algorithm.

This converter produces 280V helped voltage and fast output response is 0.009sec settling time and 0.0005sec rise time results are obtained by using a genetic algorithm (GA).

VII.   REFERENCES

[1] A. B. Wie. Jiang and B. Fahimi, "Active current sharing and source management in the fuel cell-battery hybrid power system," IEEE Trans. Ind. Electron., vol. 57, no. 2, pp. 752–761, Feb. 2010.

[2] N.Kato,K. Kurozumi, N. Susuld, and S. Muroyama, "Hybrid power supply system composed of photovoltaic and fuel-cell systems," in Proc.Int. Telecommun. Energy Conf., 2001, pp. 631–635.

[3] Taraksalmi, Mounir Bouzguenda, Adel Gastli, Ahmed "MATLAB/Simulink Based Modelling of Solar Photovoltaic Cell" International Journal of Renewable Energy Research., Vol.2,No.2, July 2011.

[4] K. Rajashekara, "Hybrid fuel-cell strategies for clean power generation,"IEEE Trans. Ind. Appl., vol. 41, no. 3, pp. 682–689, May/Jun.2005.

[5] Mona N.Eskander, "Energy flow and management of a hybrid wind/PV/fuel cell generation system," Elsvier., Energy conversion.Nov 2005.

[6] Farzam Nejabatkhah, Saeed Danyali, Seyed Hossein Hosseini, Mehran Sabahi, and Seyedabdolkhalegh Mozaffari Niapour, "Modeling and Control of a New three-input DC–DC Boost Converter for Hybrid Power System", IEEE Transactions On Power Electronics, Vol. 27, No. 5, May 2012.

[7] Noor-ul-Ain Hanif "Multiplexed control strategy for a multi-input converter using fuzzy logic algorithm", IEEE Power electronics, vol. 52, no. 15, pp. 1327–1329, July. 2016.

[8] Zhihao Li, Omer Onar, And Alirezakhaligh "Design and control of Multiple Input DC/DC converter for battery" Journal of Harvesting and Renewable Energies Laboratory (EHREL)., Vol.2, No.2, December 2011, pp.168~174.

[9] M.T. Makhloufi, Y. Abdessemed and M. S. Khireddine, "A Neural Network MPP Tracker Using a Buck- Boost DC-DC Converter for PV Systems", IEEE 5th International Conference on Systems and Control, Cadi Ayyad University, Marrakesh, Morocco, May 25-27, 2016

[10] N. Jiteurtragool, C. Wannaboon and W. San-Um, "A Power Control System in DC-DC Boost Converter Integrated with Photovoltaic using Back Propagation algorithm Artificial Neural Network", IEEE 5thInternational Conference on Knowledge and Smart Technology, 107-112, 2013.

[11] Choon-Keat Chew and Siva Ramo Rao Kondapalli "Modelling , analysis, simulation and Design Optimization (GA) of dc-dc Converter for Uninterrupitable Power supply Application IEEE, Power electronics,2005

[12] P.Suresh "Power loss optimization of boost converter using genetic algorithm (GA)"International Journal of Scientific & Engineering Research, Vol 5, issue 4, April 2014.

[13] Sathis kumar Ramoji, "Optimization of hybrid PV/Wind Energy System using Genetic algorithm (GA)," Int. Journal of engineering research and Application, vol. 4, issue 1, Jan.2014.

[14] Hamid hassanadehfrad,"Optimization of grid-connected microgrid consisting of PV/FC/UC with considered frequency control" Journal of Electrical Engineering and computer sciences, Jan 2015.

[15] A.Hina Fathima, "Optimization in microgrids with hybrid energy systems- A reviews," Renewable and sustainable energy source, Elsevier Jan 2015.

Analysis and Implementation of Text Mining for Different Documents

[1]K. Maheswari, [2]P. Packia Amutha Priya

[1]Department of Computer Applications, Kalasalingam University, Krishnankoil, Tamil Nadu, India
maheswarisnr@gmail.com[1]
[2]Department of Computer Applications, Kalasalingam University, Krishnankoil, Tamil Nadu, India
p.packiaamuthapriya@klu.ac.in[2]

ABSTRACT

The process of making structured data from unstructured and semi structured text is called text mining. Text mining is defined as bag of words. The environment is set up with various documents in a database. The preprocessing of removing unwanted numeric values, uppercase, lower case, frequent words, punctuation is considered. In this work, the frequency of words occurred at least fifty times in a document is identified. The experimental results of the word frequency in a document occurred twenty times, twenty five times, fifty times and hundred times was analyzed and represented visually.

Keywords: Text Mining, Data Mining, frequency of words and text file

I. INTRODUCTION

Text mining and text data mining [1] is a growing field of text analytics. The process of extracting the quality information from text database is known as text analytics. The quality information is extracted through analysis process. The Process of structuring input text is done by the following process,

- Parsing
- Deriving features
- The removal of punctuations numbers and will be updated into a database.

There are many databases used for implementing. They are

- Multimedia databases
- Time-series and sequence database
- Text databases
- World-Wide Web databases
- Spatial databases

A variety of information collected for processing in digital form and stores it in databases for future. They are,

- Business transactions:
- Scientific data:
- Medical and personal data
- Surveillance video and pictures:
- Satellite sensing
- Games
- Digital media
- CAD and Software engineering data
- Virtual Worlds
- Text reports and memos (e-mail messages
- The World Wide Web repositories

Today maintaining large volume of data in a database is a challenging task. The issues and challenges are developing a model of Data Mining

- Scaling Up for High Dimensional Data and High Speed Data Streams
- Mining Sequence Data and Time Series Data

- Mining Complex Knowledge from Complex Data
- Data Mining in a Network Setting
- Distributed Data Mining and Mining Multi-agent Data
- Data Mining for Biological and Environmental Problems
- Data-Mining-Process Related Problems
- Security, Privacy and Data Integrity
- Dealing with Non-static, Unbalanced and Cost-sensitive data.

A**. ISSUES AND CHALLENGES IN TEXT MINING**

There are large number of issues occur when text mining process is carried out. These issues will affect the performance of decision making. Before applying text mining process there is a need of converting unstructured data in to structured one. The issues and challenges are,

1) Challenges in Text Mining
- Bulky datasets
- Noisy unstructured data
- Ambiguous words
    Example: apple - company or apple - fruit
- Framework understanding
    Example: automobile -car -vehicle
    – Two wheeler – Four wheeler
- Composite and slight relationship   Example: "X develops software" "Software is developed by X"
- Multilingual

B. **Text Mining Process**
- Extracting Documents from the databases
- Text Transformation for processing
- Feature Extraction for analysis
- Reduce Dimensions
- Apply standard Data Mining for performance
- Interpretation / Evaluation for comparison

II. BACKGROUND STUDY

During text mining, there are a lot of issues and challenges which will definitely affect the process. The performance, efficiency effectiveness and accuracy of decision making are achieved by applying best algorithm. Before applying algorithm, pre-processing is performed. In this process, the rules and regulations are imposed to make the text process effective and efficient. This is simply defined as converting unstructured data into structured data. Variety of algorithms is used to perform text mining.

Ah-Hwee Tan [1] described a text mining structure with knowledge distillation and text refining. The author highlighted the challenges and issues of text mining. Ingo *et al.,*[2] the author surveyed the text mining facilities in R using statistical and machine learning methods. The framework was presented for text mining. The grammatical rules and context was not considered in this work.

Mustafa *et al*., [3] described the study of original and winning pattern-based method such as pattern evolving and pattern deploying to discover the hidden pattern in the text documents. Abhishek Kaushik, and Sudhanshu Naithani [4], presented the review of various text mining tools and techniques. Zhang et al [5], introduced the research position of text mining, text classification, text clustering, association rule mining.

Abhilasha [6], analysed retrieval of text data to select right method for text mining is an important task. The author also focused on automatic text mining to find effective and easy to use method. Michele *et al.,* [7] the authors developed the text mining tool for linguistic, analysed the pros and cons of the text mining tool and compared with conventional pattern classification. Zhou *et al.,* [8], suggested a new improved KNN algorithm for text classification to avoid the complexity. The results are shown with greater accuracy. Songbo Tan [9] dealt uneven text data and NWKNN algorithm was proposed. The experimental result achieves performance improvement.

III. TEXT MINING

The purpose of text classification is to increase the identification of information. The text classification needs the following,

- It obtains documents in the form of text files, pdf files, html files and other file formats.
- It contains tree structured hierarchy which specifies the important information for the

institution, company, organization and association.

- Apply software to process document using text classification algorithms.
- To obtain data from the data bases, to process it and assign to the proper classification, text mining algorithm are used.

The frame work of data flow in this work is shown in figure1. Documents are collected and preprocessed for processing. The feature extraction of documents was performed.
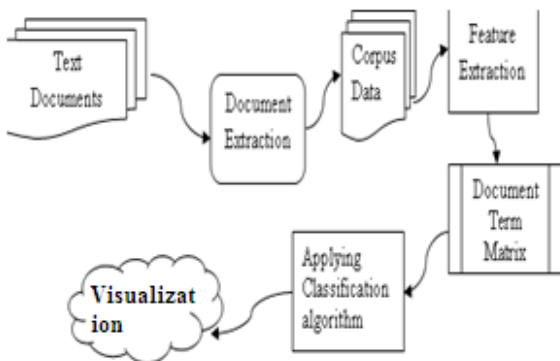
Figure 1: Frame work of proposed work

There are two approaches that are used for classification,

A. **Manual Approach**

This is the simplest approach used in text classification. It collects the key words which qualifies the type of information in a document. If more keywords are present, the key words are assigned to the topic of the document. Among the keyword the frequency of each keyword is calculated. The most frequently occurred word is assigned as topic.

B. **Statistical Approach**

This approach is based on the detection of a "training set" of the documents using data mining algorithms to find out the similarity. To deduce the key elements of the document various text classification algorithm such as Bayesian, LSA are used. In order to categorize the content, it makes use of frequency and key terms to build rules implicitly. If the classification is incorrect, there is no accuracy in result.

- **Searching of information in a database**: Data Storage and retrieval of text documents using keyword search.

- **Retrieval**: If data is present, it is extracted or retrieved from the data base.
- **Text clustering**: Text documents are grouped using data mining clustering methods.
- **Text classification**: the documents are classified using data mining classification methods based on trained dataset..
- **Web mining**: Data and text mining on the Internet is performed for finding interconnections of the web.
- **Information extraction**: The relevant data is Identified, retrieved and extracted from unstructured text;
- **Natural language processing**: The language processing is performed
- **Concept extraction**: Based on the words and phrases from the document, forming similar groups.

IV. EXPERIMENTAL RESULTS

The text files stored in the folder with the size of 1.27 MB was used in this research work. The text file with the size of 1.47 KB was considered for text mining. In this figure 2, the frequency of words in the document was found out and it is represented.
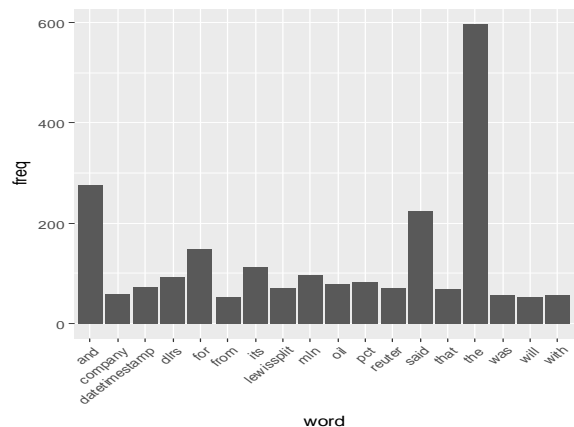
Figure 2: Word frequency of document

The most frequently used word in the document is identified and it is shown in the figure 3.
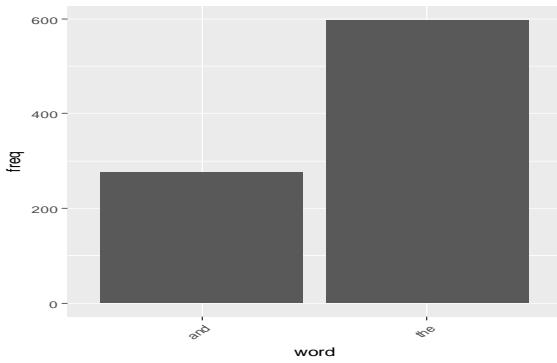
Figure 3: Most Frequent words

The identified frequent words are visually represented in word clouds and it is shown in figure 4.
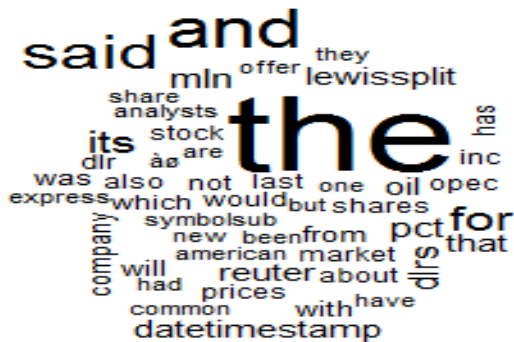
Figure 4: Cloud of Frequency word in a document

The words that occur at least twenty five times are represented visually in the figure 5.
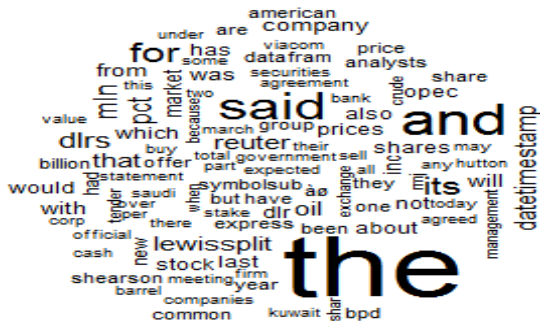
Figure 5: Cloud of Frequency word at least twenty five times in a document

The color cloud was created with the word frequent occurrence of at least twenty times is shown in figure 6.
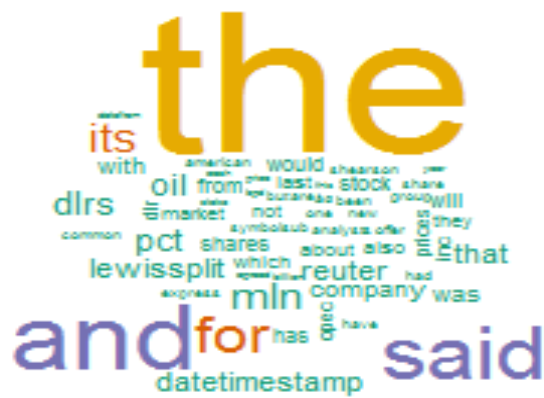
Figure 6: Color Cloud of Frequency word at least twenty times in a document

The color cloud with the word occurrence of at least hundred times in the document is shown in figure 7.

Figure 7: Color Cloud of Frequency word at least hundred times in a document

V. CONCLUSION

Text mining process is data mining is an important task in today's big data maintenance. When maintaining a huge collection of files, there are many issues and challenges are faced by day to day users. This work provides a solution for finding the frequency of words with different occurrences. The experimental results were represented visually. The future work focused for applying advanced data mining algorithms for huge volume of data.

VI. REFERENCES

[1] Ah-Hwee Tan, "Text Mining:The state of the art and the challenges", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 6, August 2012

[2] Ingo Feinerer, Kurt Hornik , David Meyer "Text Mining Infrastructure in R", Journal of Statistical Software March 2008, Volume 25, Issue 5.

[3] Mustafa M. Shaikh, Ashwini A. Pawar, Vibha B. Lahane, Pattern Discovery Text Mining for Document Classification, International Journal of Computer Applications, Volume 117 ,No. 1,May 2015,PP:6-12.

[4] Abhishek Kaushik, and Sudhanshu Naithani, "A Comprehensive Study of Text Mining Approach", IJCSNS, VOL.16 No. 2, February 2016, PP: 69 – 76.

[5] Yu Zhang, Mengdong Chen, and Lianzhong Liu, "A review on text mining", published in IEEE Xplore digital library, Software Engineering and Service Science (ICSESS), 2015 6th IEEE International Conference on 23-25 Sept. 2015.

[6] Abhilasha Singh Rathor and Dr. Pankaj Garg, "Analysis on Text Mining Techniques", IJARCSSE , Volume 6, Issue 2,February 2016, ISSN: 2277 128X, pp: 132- 137.

[7] Michele Fattoria, Giorgio Pedrazzib, and Roberta Turrab, "Text mining applied to patent mapping: a practical business case" World Patent Information, published in Elsevier, Volume 25, Issue 4, December 2003, Pages 335–342.

[8] Zhou Yong, Li Youwen and Xia Shixiong, "An Improved KNN Text Classification Algorithm Based on Clustering", JOURNAL OF COMPUTERS, VOL. 4, NO. 3, MARCH 2009, pp: 230- 237.

[9] Songbo Tan,"Neighbor-weighted K-nearest neighbor for unbalanced text corpus", Expert Systems with Applications,Volume 28, Issue 4, May 2005, Pages 667–671

A Survey of Content based Video Copy Detection using Big Data

Karthika.P*[1], P.Vidhya Saraswathi[2]

*[1]Research Scholar, Department of Computer Applications, Kalasalingam University/Krishnan koil, Tamilnadu, India
karthikasivamr@gmail.com[1]

[2]Assistant Professor, Department of Computer Applications, Kalasalingam University/ Krishnan koil, Tamilnadu, India
vidhyasaraswathi.p@gmail.com[2]

ABSTRACT

Content based video copy is the method of detecting illegally copied videos by analyzing them and comparing with original content. It extracts options from the original videos and verifies whether a duplication happens or not by looking at the extricated highlights. A state-of-the-art system of video copy detection is evaluated on VCDB to illustrate the limitations of existing techniques. The task of partial copy detection in videos aims at finding if one or more segments of a query video have (transformed) copies in a large data set. Transformations like new views are specific to 3D videos and create the copy detection even tougher. The algorithmic program realizes the 3D fake video and evaluate deep learning options learned to severally train on a distinct data sets. In this paper dynamic looking for the partial video copy detection is used to seek out a lot of segments of a query video and reference video to look the huge scale data set.

Keywords: State-of-art; Video Copy info; question video; Reference video; giant dataset

I. INTRODUCTION

Video copy detection has been actively studied in a wide range of multimedia applications. It's been described that resolution stereoscopic video less committal to writing quality and provides higher rate distortion. During the past decade there has been an exponential growth of online videos due to the huge amount of user contributed multimedia contents through abundant video sharing websites. The massive publishing and sharing of videos raises the problem of a large amount of near duplicate copies. Video copy detection is essential to many real-world applications, including copyright protection, law enforcement investigations, business intelligence, advertisement tracking and redundancy removal. Generally, a video copy is a segment of video sequence that is transformed from another one by means of inserting patterns, compression, change of gamma, decrease in quality, camcording, etc. This recompense square measure wished attributes towards meeting the wants of low bit rate applications as in handheld devices. In line with the containment theory of vision the entire perceived quality for mixed special resolution stereoscopic video is near

the read with the best quality. This is often owing to the massive frequency parts that recompense the corresponding parts within the lower special resolution frames. Uneven successive resolution and quality square measure different alternatives for uneven committal to writing. The particularly committal to writing sequences that contain quick object motion, whereas the latter produces inevitable block artifacts once committal to writing videos at low bitrates. The mixed special resolution approach provides higher professed quality than different committal to writing approaches once committed to writing multi read videos at low bitrates.

The prediction may be an essential fraction of multi read the committal to writing that exploits the temporal and cross read correlations among adjacent frames. Prediction is delineated by the organization choice and organization ordering to identify a collection of reference frames wherever they keep in the decoded image buffer. Organization ordering defines however the indices of those frames, square measure settled within the buffer list. Exponential is employed to code indices of organizations to choosing reference frames that have a most important role to put down image prediction providing an acceptable reference frame

ordering would recover committal to writing potency. This is often owing to the block matching method that the improvement of the particular bit rate and distortion through methodology. The information preparation committal to writing relationship and therefore the performance parameters utilized in the investigations reportable. The multi reads videos are used, these videos square measures canceled because the multi read committal to writing common take a look at the conditions. They cowl a good vary from multi-view videos have less distinction compared to the remaining videos since each have less put down camera distance and scene quality.

The suggestion of objects in Exit videos is slow, whereas its quick videos from the focuses on low bit rate applications the first resolution of the parts was decimated victimization the MPEG-4 filter by an element of 2 within the horizontal and vertical and diagonal directions. The ensuing videos square measure, then treated as views that contain full special resolution frames. The special resolution for frames views is more decimated so as to supply low special resolution frames. In organize to supply one stream among multi read videos, frames with totally different spatial resolutions square measure multiplexed committal to writing in order. The coded low special resolution frames, square measure interpolated victimization associate degree AVC interpolation filter coefficients for the MPEG and AVC filters. These filters, square measures prompt in uneven video committal to writing. 3 read videos are measured throughout the testing of the prediction of the context of one scene situation to come up with multi read videos with onerous scene changes and frames. The frames from every of the opposite videos that is in the center read were decimated whereas frames that the encircling views were full special resolution frames.

A successive analysis, prediction structure was victimization the experiments accessible. It permits 2 reference frames to be used for putting down image prediction. The quantization settings represent committal to writing videos acceptable were adjusted in line with the predefined values that square measure reportable within the canon takes a look at the conditions.

II. LITERATURE SURVEY

A. Anis BenHajyoussef et al [1] have presented the high efficiency video coding used to the state -of-art methods video thickness has been designed to improve coding tools and complicated techniques. A completed the original features and important coding efficiency of cost in large implementation complexity. In this complexity has increased the High Efficiency video coding encoders want for the fast algorithm and hardware implementation. The differential encoding has to perform resolution, overcoming the real time restriction while taking care of coding efficiency. The encoding complexity and preprocessing solution used for high efficiency video coding to reduce to generate the gradient and necessary approaches. The Prewitt operator used to generate and investigate particularly enhance the HEVC intra model's. A pixel based gradient preprocessing point for the HEVC intra coding algorithm uses Prewitt as a discrete differentiation operator to approximate the gradient values on the original picture. In this algorithm to generate a direction for each pixel and each production unit from select a applicable set of modes to be tested at a rate distortion optimization of through the neighbor mode selection and extension and adopted the cost function. the gradient information in order to speed up the best intra mode process on different probability to the limitation of the modes to be tested. A gradient based estimation of the texture complexity use for coding unit decision a reduction of 42.8% in encoding time with an increase in the Bjontegaard data rate of only 1.1%.

B. Pei-Yu Lin et al[2]. have presented new Augmented reality (AR) mechanism that hide the unnoticeable AR tag in the video frame supported the conception of temporal psycho-visual modulation (TPVM) and watch the naked video sometimes a 120-Hz monitor while not survey the embedded AR tag employing a mobile device the AR tag are often extracted and recognized owing to the semiconductor camera sensors capture the high frequency frames. The system is possible and applied in numerous AR application and activity. The post image process of the AR tag space like the noise elimination and therefore the optimize it thought-about to enhance the standard of the marked frames and therefore the establish the AR tag within the AR system. The associate AR tag is sometimes used and cargo on product for the sake of detected and recognized simply. The AR system recognized pattern the corresponding virtual object on the AR tag distorts the perception of the

product. The integrates the AR tag with digital screen applying the temporal psycho-visual modulation (TPVM) technique. The new methodology uses the distinction between the human eye observation and therefore the imaging of a camera to provide associate invisible AR tag is often detected by mobile devices, however not nonetheless by the human eyes support the ideas of AR and TPVM the new mechanism are smart within the connected AR application of business and amusement and protection.

C. Yinghao CAI et al [3] have to present a good spatial image and video filtering tool (GIFT) to settle on the foremost relevant input pictures and video applications to go labeled mobile videos. GIFT forcefully of mobile media glad and their Geo spatial data, video manipulate the spatial and temporal domain to traumatize giant volumes of knowledge. To demonstrate the effectiveness of GIFT we have a tendency to introduce associate degree end-to-end application that utilizes mobile videos to attain the target time. The experimental results illustrate promising performance of vision applications with GIFT in terms of lower communication lead to improved potency and accuracy and ability compared baseline approached don't absolutely utilize Geo spatial data. The Geo spatial image and video filtering tool (GIFT) to pick the foremost relevant set of input pictures and videos and applying the analysis of the GIFT persistent pursuit application was introduced and evaluated victimization each artificial and real data set. The determined a considerably improved performance of PC vision application in terms of lower communication load associate degree improved potency and accuracy to handle an outsized quantity of video knowledge. the chosen input video frames victimization property may expedite the overall time interval considerably while not impacting the accuracy of the result GIFT will be extended to serve a tool for economical video management which might be applied to media kind with geo spatial properties.

D. Imad Batioua1 et al [4] have to represent the three new divisible two dimensional distinct orthogonal moments a comparative between freelance of two dimensional distinct orthogonal moments and therefore the classical ones in terms of grey level image reconstruction accuracy as well as wheezy and noise free conditions. The native feature extraction capabilities of

the moments are a unit delineated a brand new set of RST (rotation, scaling, and translation) invariants supported divisible moments is introduced for the primary time and their description performances are a unit extremely tested as pattern options for image classification in association with the standard moment invariants. The experimental outcomes that the initial set of moments is probably helpful within the field of image analysis. The initial set of quantity distinct orthogonal polynomials supported the merchandise exploitation these quantity distinct orthogonal polynomials outlined 3 new divisible second distinct orthogonal moments named RTM, RKM, and RdHM. It measures the performance of the ways as compared to the classical well-known moments in terms of image modernization, quality native feature extraction and image classification accuracy. It highlighted that in most experiments of the moments offer higher outcomes than classical ways and their invariability is extremely confirmed. Are considering all bestowed performances and strength of this original set of moments assured of their ability to relinquish a stronger illustration of the image content which will be very useful within the fields of image analysis. It targets rising the numerical stability of the moments and presenting a quick algorithmic rule for computation of enormous size pictures rather than the simple algorithmic rule.

E. Bo-Yi Sung et al [5] in this paper accurate 3D mensuration systemsupported optical device scanners are able to acquire 3D info directly and exactly in real time. The scrutiny of the traditional cameras these forms of instrumentality are sometimes costly and thus they aren't normally on the market to customers. The optical device scanners interfere simply with one another sensors of constant sort vision primarily based 3D mensuration techniques use stereo matching to amass the camera relative position and estimate the 3D location of points on the image. It desires further estimation of the 3D info systems with real time capability typically depends on serious correspondence prevents realization on mobile devices. Impressed the structure from motion systems that reconstructs thin feature purposes to a 3D point cloud employing a mono video sequence to attain higher computation potency. The key frames to estimate the present position of the camera so as to cut back the computation load and therefore the noise interference on the system the sake of avoiding duplicate 3D purposes reconstructs the second point only the purpose shifts out

of the boundary of a camera. In experiments outcomes are in a position to be enforced is able to do progressive correctness with a denser purpose cloud with elevated speed vision community has contributed several efforts up the standard of the reconstructed 3D purpose cloud. The incontestable a system that generates associate correct purpose cloud with high speed scrutiny to 3D reconstructing ways able to use solely a mono video sequence as associate input a reconstruct the 3D purpose cloud as dense as attainable. The chiefly vital measure for a mono 3D reconstructing methodology is that the serious load on estimating the camera position will solely be guessed from the projection between frames. the strategy is in a position to lesser the load on estimating camera position, whereas behind little or no preciseness of camera baseline as a reference the calculable camera position has typically increased not solely with noise however additionally the ambiguous scale between the pixels and therefore the world. it's able to reconstruct a scenery at intervals a district additionally encounter some scale drift whereas the video sequence was verified on a really long distant hope that this sort of quick and correct 3D reconstructing algorithmic program scenery.

F. Nan Nan and Guizhong Liu et al [6] in this paper represented the frame fusion theme depends on path merging in a very graph model that is in a position to figure in an internet manner and supply informative temporary fusion results. Supported these temporary results question content predictions are often generated which can be fed back to the frame program to instruct it to adaptively modify the search strategy. The experimental results show that the projected frame fusion theme achieves competitive detection and localization accuracies compared with the progressive ways. in the meantime with the help of the adaptive search strategy the machine quality of frame similarity search is dramatically reduced at a value of a small decrease in accuracy. a unique video copy detection system that is differentiated from the present ones for the path merging algorithmic program in frame fusion and also the adaptive search strategy supported question content prediction. The experimental results show that compared with the progressive ways the projected one isn't solely able to come through competitive performance, however conjointly to scale back a considerable quantity of calculation. As a future work we might prefer to more improve the prediction

accuracy by analyzing the reference videos in advance. Additionally we are going to attempt to adopt period shot boundary detection to boost the localization preciseness.

III. PROBLEM STATEMENT

Detecting copies of a video may be a tough task the primary videos are composed of the many frames and comparison varied frames from potential video copies against reference videos is computationally intensive. Detection of video copies is additionally tough by the very fact effects occur on the traced videos. These edits typically a video transformation is done deliberately to avoid detection or accidentally owing to the repetition method. Scaling then blurring then cropping then motion picture to alter frame rate transformations. Transformations like new views are specific to 3D videos and create the copy detection even tougher. An actual copy detection system can represent the video content by strong, then discriminating and compact signatures. A discriminating cross may be a sensible quality diagrammatic of a video content, however unlikely to be found in videos that have completely different content. A compact cross needs fewer spaces for storing and fewer computations to gauge the comparison between 2 cross the search method in economical and quick to effectively confirm copies in giant databases.

IV. RESULTS AND DISCUSSION

A. The problem concentrates on in this can be stated as follows

Problem of 3D Video Copy Detection realize copies of a given 3D video n variety of videos, albeit copies square measure changed and/or embedded in alternative videos.

V. CONCLUSION

Three dimensional videos are receiving quite well-liked and making 3D videos is dear protective 3D videos adjacent to bootleg repetition is a crucial drawback. The accessible the careful style of a 3D video copy detection system has 2 main parts for one is process Reference Videos and another one is process query Videos. Within the 1st part the system creates compact cross of the depth and texture of the reference 3D videos and stores

them during an information. The second part creates similar cross for every query video and compares them against the cross within the information. If a match is found the situation of the traced half within the reference video is known and evaluated its performance in terms of exactness and racial victimization several 3D videos. These videos have 2 views the others have eight totally different views created an oversized set of question videos. The fastidiously changed the question videos to represent most sensible for repetition 3D videos specifying query videos represent the follow circumstances (i) query videos are segments of many reference videos (ii) every query video is subjected to transformations either on the feel or depth (iii) multiple collective transformations are applied on the feel and depth of every video (iv) original views are synthesized and (v) query videos have solely a separation of views of reference videos high exactness and recall values traced videos are unmodified elements of original videos and it produces over ninetieth exactness and recall traced videos to different individual transformations. The every video is subjected to 5 totally different transformations at a similar time our system yields exactness and resale values. The obtained for a good variety of the brink parameter utilized in the system doesn't would like fine standardization of that parameter.

VI. REFERENCES

[1]. Anis BenHajyoussef1,2*, Tahar Ezzedine1 and Ammar Bouallègue1 BenHajyoussef, "Gradient-based pre-processing for intra prediction in High Efficiency Video Coding" EURASIP Journal on Image and Video Processing (2017) 2017:9 DOI 10.1186/s13640-016-0159-9.

[2]. Pei-Yu Lin1*, Bin You2 and Xiaoyong Lu2 Lin, "Video exhibition with adjustable augmented reality system based on temporal psycho-visual modulation" EURASIP Journal on Image and Video Processing (2017) 2017:7 DOI 10.1186/s13640-016-0160-3.

[3]. Imad Batioua1* , Rachid Benouini1, Khalid Zenkouar1 and Hakim El Fadili2 Batioua, "Image analysis using new set of separable two-dimensional discrete orthogonal moments based on Racah polynomials" *EURASIP Journal on Image and Video Processing* (2017) 2017:20 DOI 10.1186/s13640-017-0172-7.

[4]. Bo-Yi Sung and Chang-Hong Lin* Sung and Lin, "A fast 3D scene reconstructing method using continuous video" EURASIP Journal on Image and Video Processing (2017) 2017:18 DOI 10.1186/s13640-017-0168-

[5]. Yinghao Cai1*, Ying Lu2, Seon Ho Kim2, Luciano Nocera2 and Cyrus Shahabi2 Cai, "Querying geo-tagged videos for vision applications using spatial metadata" *EURASIP Journal on Image and Video Processing* (2017) 2017:19 DOI 10.1186/s13640-017-0165-6.

[6]. Nan Nan and Guizhong Liu, "Video Copy Detection Based on Path Merging and Query Content Prediction" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 25, NO. 10, OCTOBER 2015.

[7]. P.VidhyaSaraswathi.P, Venkatesulu.M, "A Block Cipher based on Boolean Matrices using Bit level Operations"978-1-4799-4860-4/14/$31.00 copyright 2014 IEEE ICIS 2014, June 4-6, 2014, Taiyuan, China.

[8]. Vidhya Saraswathi.P, Venkatesulu .M, "A Block Cipher for Multimedia Encryption using Chaotic Maps for Key Generation " Proc. of Int. Conf. on Advances in Information Technology and Mobile Communication © Elsevier, 2013

[9]. Vidhya Saraswathi.P and M. Venkatesulu , "A Class of Boolean Matrices Possessing Inverses Under XORand AND Operations" Journal of Scientific Research, Vol.118,no.1,january 2014,pp.108-112,Scopus Indexed.

[10]. Vidhya Saraswathi.P and M. Venkatesulu, "A Secure Image Content Transmission using Discrete chaotic maps" Jokull Journal, Vol.63,No.9,pp.404-418,September-2013.

[11]. Vidhya Saraswathi.P and M. Venkatesulu, "A Block cipher for Multimedia Content Prodection with Random Substitution using Binary Tree Traversal" Journal of Computer Science, Vol.8,No.9,pp.1541-1546,August2012.

[12]. P.K.Kavitha and Vidhya Saraswathi.P, "A Survey on Medical Image Encryption" International Journal of Scientific Research in Science and Technology, ICASCT2401 | ICASCT | March-April-2017.

Safe Trek Android App using built-in Accelerometer and Magnetometer

M. Prabha[1], E. Sangita[2], B. Anitha[2]
[1]Information Technology, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India
mpr@vcet.ac.in[1]
[2] Information Technology, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India
sangita.e96@gmail.com [2]

ABSTRACT

Trekking is a hobby to some people. But it has some risk also. The main risk is that we may miss somewhere else as we don't know the direction. There are some mobile apps which provide direction using GPS. The disadvantage is that in the dense forest signal strength is weak and also GPS doesn't provide interior routes. For that we have proposed an idea that the paths we are travelling are recorded by automatically drawing line on the screen using built in accelerometer of the android phone. Instead of using network to find path, our idea is to find path by using built-in accelerometer and magnetometer. The accelerometer will record the distance covered by the user and the magnetometer will trace the directions of the user. The app will continuously record both the information and when the user needs to return to their path it will give information recorded so far. The user can easily backtrack to their path.

Keywords: *Magnetometer, Accelerometer, Trekking*

I. INTRODUCTION

The main risk is that we may miss somewhere else as we don't know the direction. There are some mobile apps which provide direction using GPS. The disadvantage is that in the dense forest signal strength is weak and also GPS doesn't provide interior routes. For that we have proposed an idea that the paths we are travelling are recorded by automatically drawing line on the screen using built in accelerometer of the android phone. In our idea without using the network we can track the path using the built-in accelerometer and magnetometer of the mobile phone. Accelerometer will record the distance travelled by the user and magnetometer will provide the direction in which the user is travelling. We can use this idea when we are travelling in the interior area of the forest or the area in which we are not having network. Accelerometer and magnetometer will continuously record the distance and the direction travelled by the user and will provide the information to them when the user returns to their path. By this, the user can easily backtrack to their path.

II. Application Design

A. Software Used

- Eclipse IDE: Eclipse is an integrated development environment (IDE) used in computer programming, and is the most widely used Java IDE. It contains a base workspace and an extensible plug-in system for customizing the environment. Eclipse is written mostly in Java and its primary use is for developing Java applications.

B. Components Used

- Android phone with Built-in Accelerometer and Magnetometer.
  1) Built-in Accelerometer: The working of accelerometer is to detect the acceleration of the phone.
  2) Built-in Magnetometer: The working of magnetometer is to find the direction.

C. Block Diagram
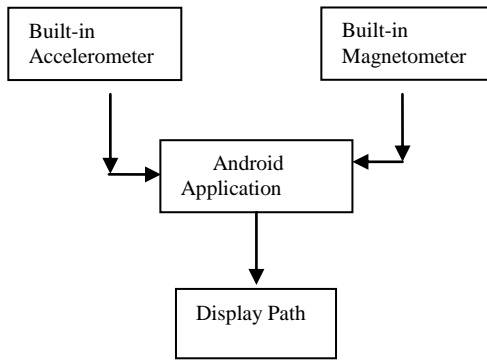
Figure 1. Block Diagram of the operation of Built-in Accelerometer and Magnetometer and the Application

III. PROPOSED SYSTEM

For the location of the path and direction we use the built-in accelerometer and magnetometer of the mobile. The accelerometer is used for measuring the acceleration of a moving body. It helps us to the find the movement of the mobile phone while we are travelling. With the movement of the mobile we can find the distance travelled so far. The magnetometer is used for finding the direction. The magnetometer of the mobile provides the direction in which we have to move.
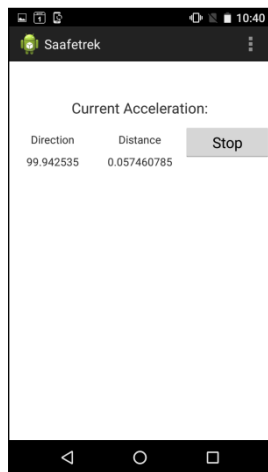
Figure 2. App showing current Direction and Distance

When we place our mobile on the flat surface, the accelerometer will show the reading of the x-axis, y-axis and z-axis as zero. The y-axis and z-axis value will be shown when we place our mobile in the upward direction. When we tilt our mobile we can get the x-axis and z-axis values.
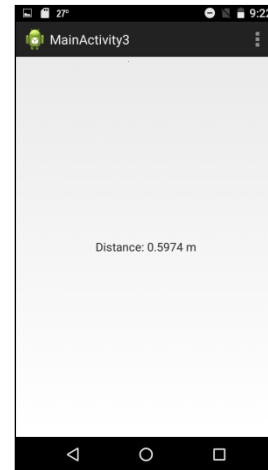
Figure 3. App showing total distance travelled.

We can measure the distance travelled so far and the direction in which we have to move when we cannot track the return path. When we are travelling we have to start our app and we can move in our path.
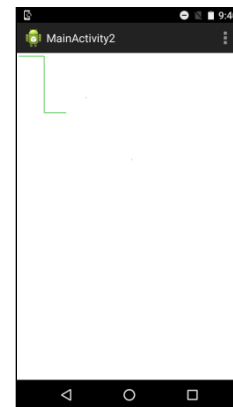
Figure 4. App showing the path travelled

When we want to return to our destination and if we don't know the path we can use the app and can track the path and return to the starting point.

IV. LITERATURE SURVEY

The application is entirely a new one and different from the app that we are using earlier. In the existing app they use the GPS of the mobile to track the path and the direction in which we have to move and they need network for tracking the path. But in our app we are using the built-in accelerometer and magnetometer of the mobile for tracking the path and the direction in which we have to move. It is helpful when we are going for trekking and when we lost our path we can use this app. So our project is different from other project in

tracking the path without the use of internet connection and if we don't have network in that area. This application can be developed into a software as it has more advantages and it is more innovative than other previous existing methods. The main difference between the existing app and this app is that we are not in the need of any network and internet connection in tracking the path and direction in which we have to move. Thus this app will be more useful in tracking the path and direction in which the user has to move.

V. CONCLUSION

This app will be an excellent method for tracking the path and the direction in which we have to move if we lost our way. It does not need any internet connection for the location of the path and the direction in which we have to move. So it is concluded that we can use in use this app in the android phone which have the built-in accelerometer and magnetometer on them.

VI. References

[1] https://eclipse.org/ide

[2] https://www.researchgate.net/publication/2204430 98_BikeNet_A_Mobile_Sensing_System_for_Cyc list_Experience_Mapping

[3] Brownlow, M.: Smartphone statistics and market share. Email marketing reports (June 2010)

[4] https://www.researchgate.net/publication/2204430 98_BikeNet_A_Mobile_Sensing_System_for_Cyc list_Experience_Mapping

Data Driven Simulation Framework for Taxi Ride Sharing

P. Potri Rathna[1], Dr.T.Revathi[2]

[1]*Information Technology, Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India*
rathnaprabhakaran@gmail.com[1]
[2]*Head of Department, Information Technology, Schlenk Engineering College, Sivakasi, Tamil Nadu, India*
trevathi@mepcoeng.ac.in [2]

ABSTRACT

In the modern era vehicles are increasing exponentially with respect to its population. The urban cities are facing many challenges in transportation and energy consumption. The foremost approach will be taxi ride-sharing which effectively reduces traffic congestion, gasoline consumption, and pollution. Our proposed method will simulate a real-time data-driven framework for analysing the taxi ride-sharing in various scenarios. In this approaches the taxies and trips are modelled as separate entities for simulating a rich set of realistic scenarios. A new optimization algorithm is described to address the computational complexity and scalability is achieved by an efficient indexing scheme combined with parallelization. The framework is evaluated using a real-time streaming information obtained from the user.

Keywords : Taxi Ride-Sharing, Shortest-Path, Scheduler, Scalability, Apache Spark

I. INTRODUCTION

Metropolitan cities are facing huge challenges due to a substantial increase in vehicles over the year, and this leads to traffic congestion, gasoline consumption, and pollution. Optimal strategy to decrease the stream of traffic and resource consumption [1] will be a taxi-ride sharing and at the same time, it needs to serve the transportation of city dwellers. The unused taxi can be efficiently filled by ride-sharing services. Every country wants to minimize its traffic and pollution, and the taxi company wants to get higher profit in each margin; and people wants to reach their destination faster with minimal cost. Sharing a taxi-ride has been identified as an optimization problem where the aim is to identify an optimal ride-sharing schedules [2], [3], [4], [5], [6], [7]. Some private organization is already providing ride-sharing services such as Uber, Ola, Lift, Via, Bandwagon and Cab With Me.

The process of deploying taxi-ride sharing needs a better understanding of its tradeoffs. The data-driven approaches are applied by providing a large volume of data for better understanding of the problem. The tradeoffs of taxi sharing are very challenging because there are multiple stakeholders with different and often with conflicting interests. Initially, the problem has been devised on the basis of survey data and analysis of psychological incentives. The graph-based model to this problem has been proposed by Santi et al. [8] which computes optimal sharing strategies for the trips and it uses two key parameters namely: maximum number of trips that can be shared and the maximum delay the passengers are willing to tolerate. This helps to study the passenger's inconvenience in sharing a taxi and for processing this model the passengers trips details need to be known in advance.

Our proposed simulation framework helps in analysing different ride-sharing scenarios. In this model, the passenger's trips are need not be known in advance and it fits well with models using different vendors. The proposed method helps in studying the realistic scenarios by providing a rich set of variables and by modelling taxi and trips as different entities which consider the different constraint of multiple stakeholders. The model includes the various variables like a maximum number of additional stops and waiting time,

and taxi-specific constraint like a number of passengers per taxi, a maximum number of shared trips. A new linear optimization algorithm is proposed for efficient indexing and assigning a trip to the taxis based on the cost factor. The efficiency of the method is analysed by providing a real-time streaming information.

II. RELATED WORK

The dynamic pickup and delivery problem [2], [3], has been addressed by linear programming [4], [5], [6]. This will be suitable for small-scale problems, for example, sharing within a certain distance. The real-time taxi dispatching uses heuristic-based method [12] but the scalability is very limited.

The data-driven method evolves with larger benefit and flexibility and the proposed method is similar to Santi et al. [8]. They used a "shareability network", where a node represents taxi trips, and the nodes get connected if they share the trips. This shareability network depends on two parameters the maximum number of shared trips per service and the maximum delay a passenger can tolerate in sharing a taxi. Let k be the maximum trip it can share and Δ be the maximum delay. This problem may lead to NP-hard for the higher k value and, it is tractable only for $k=2$. Similarly, the network size increases with Δ value, for larger Δ value the network will be larger and this will increase the computation time. However, the obtained results from the model will not be feasible for real time scenarios because this will not explicitly consider the taxi positions and their capacity

For example, there are two trips t_1, t_2 and this trips needs to be shared in a beneficial manner. For serving the passenger request there should be a cab c. Consider that c is serving some trips t_1 and some other trip t_i. If t_1 and t_3 are served by c in this order then it can't serve to the request of t_2 and similarly, if t_1 and t_2 are assigned to c then it can't serve to t_3 anymore or if it serves then the cost will be more.

According to this method, the trips of the passenger has to be known in advance for processing, this will be another limitation of this. This method is well suited for car pooling because in car pooling the time and location of each trip are fixed but this assumption is not suitable for taxi ride-sharing where trip request arrives at dynamic and real. To solve this issue Santi et al.

proposed a refinement technique to their model which prunes the shareability network by keeping a time window δ. However, in real time, this works only for $\delta=0$.

Ma et al. [9], [10] proposed a taxi ride-sharing dispatching method in real-time. Ma et al. splits the region into grid cells and calculate the distance "heuristically". This attains the fastest response because the heuristic calculation provides the shortest path with minimal computation. The accuracy of the system is very less and the results depend on the selected grid size. Similar to Ma et al. the proposed method response time will also be fast if the queries match with the cabs. Cache coherent indexing scheme is used for finding exact shortest path scheme for optimization.

Hung et al. [11] scheduling algorithms match the taxi with dynamic passenger trip requests. The taxis are scheduled with minimal cost by satisfying trip waiting and service time constraint. This uses Kinetic tree algorithms like branch-and-bound and mixed-integer programming. This type of kinetic algorithms can be integrated with the proposed algorithm.

III. DATA DRIVEN SIMULATION

When a dynamic trip request is issued by the passenger, our model schedule a taxi for the trip request by optimizing the pre-defined cost function under the constraint.

3.1 Simulation Components

Taxi Fleet: Indicates the set of taxis that are participating in taxi ride-sharing and the taxis are on dynamic. Here taxis are considered as a distinct entity which includes parameters like a maximum number of shared trips, passenger capacity, maximum pickup waiting time, extra time for drop off, vehicle speed, occupancy of people, the list of stops for sharing.

Passengers: People involve in sharing and the passenger size should be more than or equal to one. Includes the parameters like drop off location and a set of sharing constraint.

Scheduler: The scheduler schedules the appropriate taxi for the request based on ride sharing constraint. The scheduler must aware of taxis location and occupancy status.

Road Network: A directed graph G(V,E) is used to represent the road network. when a vehicle flows in both direction of a road then the bi-directed graph is used. Weights are used to incorporate the traffic condition.
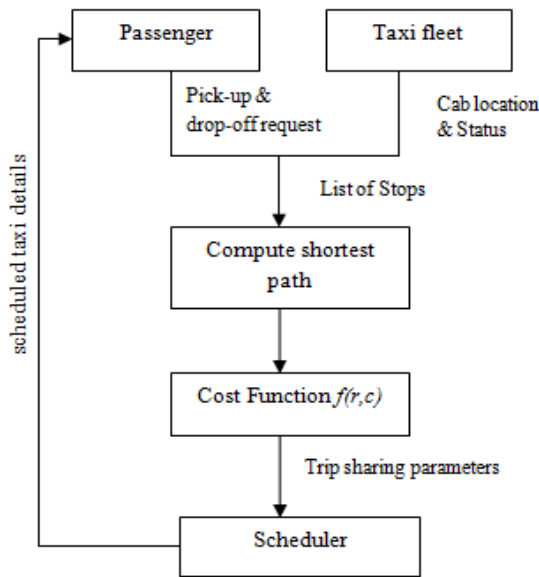
Figure 1: Overview of the Taxi Ride-Sharing

3.2 Data-Driven Simulation

The ride-sharing scenarios are simulated based on a set of input parameters and operate in an event-driven fashion where the states are updated when a new pickup request has been issued. When a new pickup request issued by the passenger then all taxis needs to update their status to the scheduler. Then the scheduler picks an appropriate taxi with minimal cost by computing the additional cost function *f*.

Input Parameters	Description
m	number of taxies involving in ride-sharing
C	Default taxi capacity
n_{share}	Maximum number f trips to be shared
t_{delay}/d_{delay}	Additional time/distance delay each customer could spend by default
t_{extra}/d_{extra}	Extra time /distance each customer could spend by default
$f(r,c)$	A cost function that a given taxi C and a pick-up request r, returns the cost of accommodating r with C.

Table 1: Input parameters for simulation

IV. SIMULATION ALGORITHM

Figure 1 shows the architecture of the proposed method. The aim is to minimize the total cost or maximize the total utility while meeting a set constraint. Let the additional cost function be $f(r_i, c_j)$ for a cab c_j for the trip r_i, n be the number of trips and m be the number of taxis. The total travel cost T(i) for the initial i trips can be calculated using

$$T(i) = \begin{cases} T(i-1) + min_{1 \le m}\{f(ri, cj)\} \\ T(i-1), if\ there\ is\ no\ available\ cabs \end{cases}$$

All the trips are considered in a chronological order. The real time dispatching is done by minimizing *Tc(n)*. The cab detail c_j is updated for every trip t_i based on the elapsed time.

4.1 Simulation Phase

The simulation takes the data in an event-driven fashion for scheduling the trip with a taxi of minimal cost with respect to set of constraint. The parameters are shown in table 1.

_____

Algorithm 1 Simulating taxi ride-share
_____

Input: a set of trips, R={r1,r2}

Parameters : m-no.of c abs, c-capacities of cabs, $n_{share,}$ $n_{delay,}$ $n_{extra,}$ f(r,c)

STEPS :

1. Sort the values of R in chronological order and store it in R

2. Initialize the capacity to all cabs C {c1,c2,…., c_m }

3. Iterate i till |R| by assigning i value as 1

4. Subtract TimeofPickUp(r_{i-1}) from TimeofPickup(r_i) and store it in ElapsedTime

5. Assign infinity(∞) to f*

6. Iterate j till m by assigning j value as 1

7. updatecabstatus (c_i)

8. Assign f($r_{i,cj}$) to f_{ij}

9. if f* greater than f_{ij} then

10. assign f_{ij} to f*

11. assign c_j to c*

12. end if

13. end for

14. Assign trip r_i to the cab c* (r_i ,c*)

15. Add T(i-1) and f* and store the value in T(i)

16. end for

As of Algorithm 1, each trip is considered in chronological order. For each passenger pickup request t_i, examine the cab C status and update the status based on their elapsed time. Additional cost $f(t_i, c_j)$ for each cab is calculated and assign ri to the cab with minimal additional cost. The speed of the cab is cab is needed for updating its state hence, calculate speed using the trip duration and distance.

Simultaneously, update the cab occupancy and planned stops.

_____

Algorithm 2 Cost Function $f(t,c)$

Input: r: trips, c:cab

Parameters : n_{share} , n_{delay} , n_{extra},

STEPS :

1. Assign list of stops of C including its current location $\{s_0, s_1, \ldots, s_k\}$ to S
2. Calculate Shortest Path between s_k and r.p_{drop} and store the value in D_{drop}
3. Compute FindPickUpLocation(r,c,s , D_{drop}) and store the value in idx_{pick} ,D*
4. Check if idx_{pick} is less than or equal to k then
5. subtract D_{drop} from D* and store in D_{pick}
6. Compute FindDropOffLocation (r,c,s ,idx_{pick} ,D*,D_{pick}) and store the value in idx_{drop} ,D*
7. end if
8. return D*

Output D* : an additional distance for cab c to accommodate

The Additional cost calculated in Algorithm 2, is used for finding the optimal route for the cab c_j which includes the pick-up and drop-off locations of r_i and compare its cost with the current route for c_j. This optimal route computation is called as Sequential Ordering Problem (SOP) which is a kind of Travelling Salesman Problem [18]. The heuristic search is used for computing the best route for c_j, for this first find the pick-up location p_{pick} and place it in the stop S and assume that the order to visit the stops are same and similarly, the drop-off p_{drop} is added to end of the route. While computing additional cost $f(r_i, c_j)$, the occupancy of the taxi is also checked. Once r_i to the cab c_j is assigned with the minimal cost, update the stops S details . The list of scheduled stops of cab c_j will be S=$\{s_0, s_1, \ldots, s_k\}$, and p_{pick} and p_{drop} be the pick-up and drop-off locations of r_i.

Let us assume that the drop-off will happen after the last stop s_k for computing the additional distance to accommodate r_i Insert p_{pick} between s_{l-1} and s_l. If D_1, D_2 and D_3 be the lengths of shortest paths between s_{l-1} and p_{pick}, p_{pick} and s_l, and s_{l-1} and s_l then the additional distance will be defined as, D= $D_1 + D_2 - D_3 + D_{drop}$, and the algorithm performs pruning and stops if the delay constraint is no longer satisfied (Algorithm 3) by using this algorithm pick-up order of passenger is calculated.

_____

Algorithm 3 FindPickUpLocation

Input: r trip, c cabs, s=$\{s_0, s_1, \ldots, s_k\}$ list of stops of c including its current location, D_{drop} the shortest distance between s_k and p_{drop}

Fields: c.C capacity of c, r.o_p number of passengers of r

Parameters: n_{share}, n_{delay}, n_{extra}

STEPS:

1. Assign the cab passenger count value to idx
2. Assign infinity(∞) to newdist(D*)
3. Increment the k value and store it in idx_{pick}
4. Iterate the for loop till k, by assigning idx+1 value to j.
5. Compute the shortest path between previous location and current pick-up location then store the value in D_1
6. Compute the shortest path between current pick-up location and stop$_j$ location then store the value in D_2.
7. Compute the shortest path between previous location and stop$_j$ location then store the location in D_3.
8. If $d_{delay} \leq 1$ && $d_{extra} \leq 2$ for all trip then
9. D=$D_1 + D_2 - D_3 + D_{drop}$
10. Check if olddist(D) < newdist(D*) then
11. Assign D value to D*
12. Assign j value to idx_{pick}
13. end if
14. else if d_{delay} not satisfied for trip r
15. break
16. end if
17. assign s_j value to prev
18. end for

Return idx_{pick}. D*

Similarly, the drop-off of the passenger is computed using the Algorithm 4. In the algorithm, the best position for p_{drop} is searched. As like the FindPickUpLocation process, for each $s_l \in \{s_{p-1} = p_{pick}, s_p, s_{p+1}, \ldots s_{k-1}\}$. Then query the shortest path between s_l and p_{drop}, p_{drop} and s_{l+1}, and s_l and s_{l+1}, will be D_4, D_5, D_6 respectively. Then the additional cost of each new route will be defined as , D = $D_{pick} + D_4 + D_5 - D_6$. For example, consider $l=5000$, $M=30$ and $\epsilon = 0.2$, the probability that the minimal cost derived by the heuristic would be more than 0.2.

_____
Algorithm 4 FindDropOffLocation

Input: r trip, c cabs, s={s_0, s_1,….., s_k} list of stops of c, idx_{pick} index of S where to insert p_{pick}, D* additional distance obtained from FindPickUpLocation, D_{pick} additional distance to insert p_{pick} at idx_{pick}

Parameters : $d_{delay,}$ d_{extra}

STEPS:

1. Assign p_{pick} value to prev
2. Iterate the for loop till k, by assigning idx_{pick} value to j
3. Compute the shortest path between the previous location and passenger drop location then store the value in D_4
4. Compute the shortest path between passenger drop location and stop(s_j) location then store the value in D_5.
5. Compute the shortest path between previous location and stop(s_j) location then store the location in D_6.
6. Check if d_{delay} <= 1 && d_{extra} <= 2 for trips are satisfied then
7. $D=D_{pick} + D_4 + D_5 - D_6$
8. If olddist(D) is less than newdist(D*) then
9. Assign D value to D*
10. Assign j value to idx_{drop}
11. end if
12. end if
13. assign s_j value to prev
14. end for

Return idx_{drop}. D*

4.2 Cache Coherent Shortest Path Index

The shortest path queries are used extensively in our algorithm (Algorithm 2, 3, 4), at this point the computation spends much time. Each computation of $f(r_i, c_j)$ makes a series of the queries for all stops of c_j, for the finding solution with minimal cost. Initially, precompute all the shortest distance between each nodes and cache the distance for all intersection pairs. This will reduce the cache misses and can be used for fast retrieval of distance if the same queries repeated. The matrix storage size is simply small and would fit completely on commodity computers. Thus the shortest path queries are now reduced to memory access.

An efficient and easy method to increase cache coherent of shortest path lookups has been proposed. Transport the shortest path matrix, forward lookups become backward lookups and vice versa and an additional transposed matrix has been included in the shortest path queries to convert all forward lookups to backward lookups. The cache-coherent layout will systematically

reduce the number of cache misses. This efficiently reduces up to six times than the single core.

V. EXPERIMENTAL SETUP

The experiment is performed with real-time user data. For getting the user information the android application is used and for executing the process Apache Spark is used. The streamed dynamic data is processed continuously using Spark.
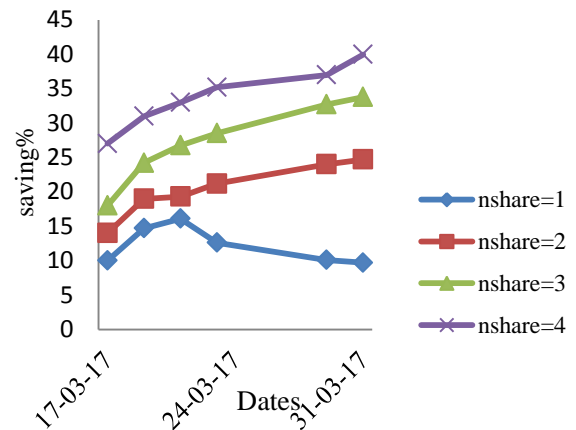
Figure 2. The saving percentage of total cost for travel distance through sharing ride for n_{share}=1,2,3,4.

5.1 Data

An Android application is created for getting user details which include information like a passenger name, number of passengers for a ride, pick-up and drop-off location and their date and time of travel. The application is created using Android Studio IDE. The application will use the Google Map detail for getting the user pick-up and drop-off location. In this, the android application will act as a client end which sends data to the server.

The Apache Spark will act as a back end for processing. It seamlessly gets user data dynamically and runs the scheduling algorithm for scheduling the user request with an appropriate cab. For processing the user request each trip is represented with the following fields: taxi ID, pick-up time, pick-up and drop-off locations, travel distance in kilometers and number of passengers details are stored in server side.   Assume that the maximum number of extra stops, which is at most $2n_{share}$. For keeping the waiting and service times as reasonable, set

d_{delay}=5 minutes and d_{extra}= 10 minutes and set each taxi's capacity C=4 for simplicity. The Fig.2 explains the total cost saving % for a passengers who undergone for sharing taxis where n_{share}=1,2,3,4

VI. CONCLUSION

In this paper, we present a data-driven simulation for taxi-ride sharing to improve flexibility and scalability for rich set of realistic ride-sharing scenarios. The main goal of scheduling algorithm is to work seamlessly in allocating a taxi request with a cab of minimal cost Cache-coherent layout helps to speed up the shortest path queries and speed up the entire processing. The implementation of our model if fully done with Apache Spark, which enables of variety batch analysis tasks. The future work we would like to implement the load balancer for the shortest path queries. In this, the shortest path database could be loaded on a separate machine. This will allow making a better use of computing resources when having multiple simulator instances.

VII. REFERENCES

[1]. V. Handke and H. Jonuschat, Flexible Ridesharing. Springer Berlin Heidelberg, 2013.

[2]. J. Yang, P. Jaillet, and H. Mahmassani, "Real-time multivehicle truckload pickup and delivery problems," Transportation Science, vol. 38, no. 2, pp. 135–148, 2004.

[3]. G. Berbeglia, J.-F. Cordeau, and G. Laporte, "Dynamic pickup and delivery problems," EJOR, vol. 202, no. 1, pp. 8 – 15, 2010.

[4]. A. Marin, "Airport management: taxi planning," Annals of Operations Research, vol. 143, no. 1, pp. 191–202, 2006.

[5]. G. Keith, A. Richards, and S. Sharma, "Optimization of taxiway routing and runway scheduling," in AIAA Guidance, Navigation and Control Conference and Exhibit, 2008.

[6]. P. Shaw, "Using constraint programming and local search methods to solve vehicle routing problems," in Principles and Practice of Constraint Programming CP98, ser. Lecture Notes in Computer Science, M. Maher and J.-F. Puget, Eds. Springer Berlin Heidelberg, 1998, vol. 1520, pp. 417–431.

[7]. B. Coltin and M. Veloso, "Scheduling for transfers in pickup and delivery problems with very large neighborhood search," in AAAI, 2014, pp. 2250–2256.

[8]. P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. Strogatz, and C. Ratti, "Quantifying the benefits of vehicle pooling with shareability networks," PNAS, vol. 111, no. 37, pp. 13 290–13 294, 2014.

[9]. S. Ma, Y. Zheng, and O. Wolfson, "T-share: A large-scale dynamic taxi ridesharing service," in ICDE. IEEE, 2013.

[10]. S. Ma, O. Wolfson, and Y. Zheng, "Real-time city-scale taxi ridesharing," IEEE Transactions on Knowledge Discovery and Data Engineering, vol. 27, pp. 1782–1795, 2015.

[11]. Y. Huang, F. Bastani, R. Jin, and X. S. Wang, "Large scale real-time ridesharing with service guarantee on road networks," PVLDB, vol. 7, no. 14, pp. 2017–2028, 2014.

[12]. W. M. Herbawi and M. Weber, "A genetic and insertion heuristic algorithm for solving the dynamic ridematching problem with time windows," in GECCO '12, 2012, pp. 385–392.

Data Mining Techniques used in Climate Analysis - A Review

N. Krishnaveni*[1], Dr. A. Padma[2]
*[1] Department of Computer Science and Engineering,P.S.R Engineering College,Sivaksi,Tamil Nadu, India
veniengg@gmail.com[1]
[2] Department of Computer Science and Engineering, Muthayammal Engineering College, Rasipuram Tamil Nadu, India
giri.padma2000@gmail.com[2]

ABSTRACT

Data mining is technology is popular innovation. It converts files of data into useful information and knowledge that can help the data originators/users to pick choices and decide smart actions for data originators benefit. Data mining is the process to find for hidden patterns amongst vast sets of data. This data is useful to understand, predict behaviors for future. Overall, Data Mining is the set of methodologies used in analyzing data from various dimensions and perspectives, finding previously unknown hidden patterns, classifying and grouping the data and summarizing the identified relationships.

Keywords: Data mining, Clustering, Classification, Big data, R tool.

I. INTRODUCTION

Climate data analysis is an active research area, focused on analysis of change of climate conditions, analysis of climate phenomena, and analysis of interconnections of climate conditions. Data mining techniques provide better and faster analysis of large amounts of data in climatology. Climate data analysis, performed in order to understand climate change process and effect of different environmental factors in that change, has been focus of interest of researches for many years. The information extracted from database is used for the development of information technology. It generates large amount of data and huge volume of data in variety of research fields. To do research in variety of fields, data mining has given a various technology to store data and operate formerly stored data and to make decision based on mining process.

There are two methods to predict weather
1. **Empirical Approach:** This approach depends on investigation of past chronicled information of forecast which is gathered in meteorologist's focus and its connection to an collection of environmental variables over various parts of areas. The mostly utilized methodologies for climate predictions are Regression, decision tree, artificial neural network, fuzzy logic and LDA.

2. Dynamical Approach: In this approach, expectations are formed by physical models.By using these models, an arrangement of conditions that predict the future climate outline. To anticipate the climate by numeric means, meteorologist has generated atmosphere models that approximate the adjustment in temperature, weight. Climate speculation expectation is actualized. It is used with the utilization of exact measurable method.

II. DATAMINING PROCESS

Data mining is used to dig out inherent and earlier unknown information from data. Data Mining is the process which affords a concept to magnetize attention of users. Availability of massive amount of data, that needs to convert into useful information. So, many people bring into play the idiom "knowledge discovery device"or KDD for data mining. Seven sequential steps are used in data mining. [1]They are

1. Data cleaning: This phase removes noise and inconsistent data from collected raw data.
2. Data integration: This phase groups multiple data sources into single data store. It is called target data.

3. Data Selection: In this, data significant to examine task are get backed from data base as pre-processed data.
4. Data transformation: In this step, data is changed or merged into typical forms for mining by performing summing up and aggregation functions.
5. Data Mining: A range of smart techniques are applied in order to dig out data patterns.
6. Pattern evaluation: Data patterns are assessed.
7. Knowledge presentation: In this stage information is symbolized using representation techniques.

The purpose of knowledge finding and data mining process is to discover the patterns that are buried among the huge set of data and understand constructive knowledge and information.

III. PREDICTION TECHNIQUES

1. Artificial Neural Networks (ANN)

ANN is a mathematical model based on biological neural networks. It consists of an organized group of artificial neurons. It processes information using a connectionist move toward to calculation. Layers of Neurons are structured. The input layer consists of the original data, while the output layer represents the classes. There may be several hidden layers. Iterative learning process is the main feature of neural networks. In which data samples are offered to the network one at a time, and the weights are adjusted to foretell the correct class label. Advantages of neural networks include their high tolerance to noisy data. Ability to catalog patterns on which they have not been trained. Main concern of the training phase is to focus on the inner weights of the neural network, which is used according to the transactions used in the learning process. For each training transaction, the network receives in accumulation the expected output.

2. Bayesian Classifier

It is a, statistical classification approach based on the Bayes theorem.

Theorem:

To estimate probability of A given B,

P (B given A) = P (A and B)/P (A) the algorithm calculates the number of cases where A and B occurs concurrently and segregates it by the number of cases where A alone happens. Let X be a data tuple, X is judged Let H be some hypothesis, such that the data tuple X belongs to class C. P (H|X) is posterior probability, of H conditioned on X.P (H) is the prior probability of H in contract.

3. Decision Tree

It makes use of the simple divide-and conquer algorithm. In these tree structures, leaves represent Classes and branches indicate conjunctions of features. It leads to those classes. The attribute most efficiently splits samples into different classes. A path to a leaf from the root is found depending on the assessment of the predicate at each node that is visited. To see coming the class label of an input. Decision tree is fast and easy method. It does not want any domain information. In the decision tree inputs are divided into two or more groups continue the steps till to complete the tree.[]

4. Hidden Markov Models

HMM is double implanted much more complex stochastic process. It provides mechanisms to teach the Markov Model <S, T> underlying a POMM from the sequence of clarification. Baum-Welch algorithm learns transition and observation probabilities. And also the state space (only the number of states has to be given). It is used to explain the observed training sequences. [10]

IV. DATA MINING TOOLS

1. Rapid Miner (erstwhile YALE)

Written in Java, gives advanced analytics methods. It is very popular because it is a readymade, open source and no-coding required. It incorporates multidimensional data mining functions such as data preprocessing, visualization, predictive analysis. It can be easily integrated with WEKA and R-tool to directly give models.

2. WEKA

It is a free JAVA based tool. It includes techniques for visualization, analysis, modeling, clustering, association, regression and classification.

3. R-Programming Tool

R programming is written in C. It permits the data miners to write scripts .So, it is utilized as to formulate statistical and analytical software. Graphical analysis,

classification, and clustering and time-based data analysis are maintained by R.

4. Python based Orange and NTLK:

Due to ease of use and its powerful features Python is very popular. Orange, NTLK are the Python based open source tool that are very powerful language processing data mining tool .They are useful for data analytics, text analysis, and machine-learning and data scraping features. These features can easily be built up for adapted requirements.

5. Knime:

Mainly used for data preprocessing. Knime is a powerful tool with GUI that gives you an idea about the arrangement of nodes. Financial data analysts is a popular one and it has modular data pipe lining, machine learning, and data mining perceptions freely for building business intelligence reports.

V. WEATHER DATA MINING

The increasing research area in data mining technology is weather data mining. Data mining holds immense promising for weather forecasting to allow climate system to systematically use data and analysis to progress. In the weather forecasting managing data mining prediction are playing vigorous role. Some of the prediction based data mining techniques are as follows:

1. Artificial Neural network
2. Bayesian Classifiers
3. Decision tree and clustering.
4. Support Vector Machine (SVM)

In this system, R-Programming is used to analyze the climate data and used to predict weather.

VI. HADOOP AND MAP REDUCE PROGRAMING MODEL

[11] Hadoop and Map Reduce framework are the most widely used models today for Big Data processing. Hadoop is an open source large-scale data processing framework. It supports distributed processing of large chunks of data using simple programming models. The

Apache Hadoop project consists of the HDFS file system and Hadoop Mapper and Reducer function. Hadoop is an open-source framework for dealing out a large amount of data across clusters of computers with the use of high-level languages like Java. Its modules provide easy to use languages, graphical interfaces and administration tools for managing data. Hadoop cluster is a set of machines networked together in one location. Data storage and processing all occur within this cloud of machines. User can submit jobs to Hadoop from his desktop machine in remote location from the Hadoop cluster[5].Two main components of Hadoop are Hadoop Distributed File System (HDFS) and MapReduce[11]. HDFS is a distributed file system management for large datasets of sizes of gigabytes and petabytes. And MapReduce is a programming framework for managing and processing huge amounts of unstructured data. Parallel processing of big dataset into smaller independent chunks.

VII. BIG DATA USING R

R is an open source software platform. Scope of big data analysis using R for statistical data analysis is popular. R is promptly adopted by statistics departments in universities Comprehensive R Archive Network and make it available to everyone. An excellent open – source interactive development environment has been created by R Studio for the R language. To increase the productivity of R users, [5] Google, Ford, Twit ter, US National Weather Service, The Rockefeller Institute of Government, The Human Rights Data Analysis Group makes use of R. Around the world, fascinated by its extensible nature as a platform for academic research. [5]Free in cost surely played a role as well. And it wasn't long before researchers in data science, statistics and machine learning started to publish papers in academic journals along with R code applying their new methods. R builds this process very easily and anyone can produce an R package to CRAN that stands for Comprehensive R Archive Network and make it available to everyone. An excellent open-source interactive development environment has been created by R Studio for the R language, further boosting the productivity of R users everywhere. [1]Google, Ford, Twitter, US National Weather Service, The Rockefeller Institute of Government, The Human Rights Data Analysis Group makes use of R.

VIII. CONCLUSION

To create a powerful and reliable statistical model, the following processes are more important. Like data transformation, evaluation of multiple model options and visualizing the results are essential. This is the reason why the R language has proven so popular: its interactive language uplifts investigation, explanation and presentation. Revolution R Enterprise gives the big - data support and speed to allow the data scientist to repeat through this process quickly. R programming model is used to analyze the climate data in very efficient manner.

IX. REFERENCES

[1] Joshi A, Kamble B, Joshi V, Kajale K, Dhange N. Weather forecasting and climate changing using data mining appli¬cation. International Journal of Advanced Research in Computer and Communication Engineering. 2015 Mar; 4(3) : 19–21.

[2] Caesar Wu, RajkumarBuyya, and Kotagiri Ramamohanarao. "Big data analytics = machine learning + cloud computing". CoRR, abs/1601.03115, 2016. Kaur, Anureet. "Big Data: A Review of Challenges, Tools and Techniques. IJSRSET, 2(2), 2016.

[3] Ijarcce Issn (Online) 2278-1021 Issn (Print) 2319 5940 International Journal Of Advanced Research In Computer And Communication Engineering Vol. 5, Issue 6, June 2016 Copyright

[4] To Ijarcce Doi 10.17148/Ijarcce.2016.56142 643 Weather Prediction Based On Big Data Using Hadoop Map Reduce Technique Basvanth Reddy 1, Prof. B.A Patil 2

[5] Big data analytics using r sanchita pati International research journal of engineering and technology (irjet) e-issn: 2395 -0056,volume:3 issue: 07 | july-2016www.irjet.netp-issn: 2395-0072

[6] Olaiya F. Application of data mining techniques in weather prediction and climate change studies. I. J. Information Engineering and Electronic Business. 2012 Jul ; 1: 51–59.

[7] Kalyankar MA, Alaspurkar SJ. Data mining technique to analyse the metrological data. International Journal of Advanced Research in Computer Science and Software Engineering. 2013 Feb; 3(2):114–118.

[8] Chauhan D, Thakur J. Data mining techniques for weather prediction: A review. International Journal on Recent and Innovation Trends in Computing and Communication. 2014 Aug; 2(4):2184–2189.

[9] Han, J., Micheline K., 2007, Data Mining: Concepts and Techniques, San Fransisco, CA: Morgan Kaufmann publishers.

[10] Rohit Kumar Yadav,Ravi Khatri 2016,"A weather Forecasting Model using the DataMining Technique,International Journal of Computer Application,Volume 139-No .14

[11] Dr. Doreswamy,and Ibrahim GadBig Data Techniques: Hadoop And Map Reduce For Weather Forecasting, International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2016, pp. 194-199 e-ISSN:2278-621X.

[12] Basvanth Reddy1, Prof. B.A Patil2 Weather Prediction Based on Big Data Using Hadoop Map Reduce Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 6, June 2016

[13] http://www.r-statistics.com/tag/hadley wickham/

[14] www.tutorialspoint.com

A Novel Approach to Evaluate the Service Quality by Exploring Social User Contextual Information

E.Jai Vinitha[1], J.Maruthu Pandi[2], R.Divya Bharathi[3]

[1]Department of IT, M.Tech, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
evinithait12@gmail.com[1]
[2]Department of IT, Assistant Professor, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
jmaruthupandi@mepco.ac.in[2]
[3]Department of IT, M.Tech, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
bharathikrishnan2.db@gmail.com[3]

ABSTRACT

With the increase of social media and e-commerce, enormous people prefer to share their experience and rate on review websites. Existing research are mainly focused on personalized recommendation and rating prediction but evaluating the quality of service for recommender system is more important. The proposed approach focuses on service quality evaluation. There are some challenges that do not have enough review information for extracting opinion. In this paper, a Service Quality Evaluation model is proposed to evaluate the service quality. The proposed model can be done in three steps. First step is to calculate the entropy which is utilized in users' confidence value. Second, to explore the contextual features of user rating in which the spatial-temporal features and sentimental features are reviewed. The final step is to fuse the above two steps into a unified model for calculating the overall confidence value to perform service quality evaluation. The experiments are implemented by using Yelp and Douban dataset.

Keywords : Spatio-temporal features, sentimental features, Data Mining, Contextual Information of User

I. INTRODUCTION

Nowadays, with the development of mobile devices and internet access, social network services have become popular. Users share their experiences like movie, ratings, and moods on internet. The first generation of recommender system [2]-[12] and social network based models [13]-[22] mainly focus on personalized recommendation and predicting users preferences but they all ignore the service quality. Thus, we mainly focus on quality of service.

When we choose an item, we mainly rely on users review and rating. Generally rating ranges from 1 to 5, the more user rating to the item there are more confidence in the overall rating. For example, consider for a movie the given rating is 4.5 by hundreds users we will assume that the movie is good. However, there will be some users who will not like the movie and the rating

is two. The audiences will get confuse if the rating are contrary for same item. If the rating is contradiction then we take average for the rating and consider it as overall rating.

There are several challenges for service quality. The first challenge is the rating sparsity. The second challenge is user confidence bias. Users have different pattern for services. The third challenge is user confidence which is not isolated. In addition users may give high rating but there may be several negative reviews. So, we need to explore users' rating confidence by closely examining social users' contextual information.

In the proposed method we first utilize information entropy to calculate user rating confidence. Second from the users' contextual information the spatial-temporal and sentimental features of rating are examined. Finally,

they are fused together to form the unified model. The main contribution of the paper is as follows:

- The issue of quality evaluation for service is addressed and the probabilistic linear model is proposed for exploring users' contextual information.
- We use the user rating confidence to evaluate the quality of service because different users have different level of confidence. Further user profiles are changing at different places in different times. So, we implement probabilistic linear model with Gaussian observations.
- We find contextual information for constraining user rating confidence. User rating is higher when a user is very far away from the rated item.

The remainder of the paper is organized as follows. In Section 2 related work on recommender system. In Section 3 proposed model is presented. In section 4 the description about dataset is given. In Section 5 experiments are presented. Finally in Section 6 conclusion is explained.

II. RELATED WORKS

Koren [29] proposed temporal dynamics with the collaborative filtering and he considered only user and item time which changes and compared with various baselines.

Dror et al.[30] proposed a model that captures information from taxonomy of items and different temporal dynamics of music ratings and the idea can be used in user bias which convert the personalized rating prediction to service quality.

The matrix factorization model [6], [7], [21]-[27], [29]-[33] predicts the user's ratings, in which the unknown ratings are predicted by using the latent features of users and items. In the previous work [21], [22], [25], [26], [27] considers the social factors in matrix factorization, including interpersonal influence, interest similarity, personal interest etc.

Multimedia recommendation is addressed in [19], [20], [34], [35], [36]. Lee et al. [34] proposed the recommendation concepts for both novel and relevant recommendations. Wang et al. [19] proposed a

framework for suggesting the videos that users import in the online social network.

Existing works focus on personalized recommendation or rating prediction but we focus on service quality evaluation by exploring social users' contextual information. For predicting the users' rating matrix factorization can also be used [23], [24], [25], [26], [27], [37]. Sarwar et al. [2] proposed an algorithm for item collaborative filtering in which they focus on predicting the users' rating of an item by calculating the average ratings of similar items.

The sentiment analysis method [38], [39], [40] focus on social networks, public sentiment, and web queries. Zhang et al. [38] proposed self-supervised emotion-integrated sentiment classification results into collaborative filtering in which user-item rating matrix is inferred by decomposing item reviews that user give for item.

Tan et al. [39] proposed a model for text collection in comparison with another background text collection. The cold start problem is the item with only few ratings and research is focused on cold start problem [41]-[43]. Leroy et al. [41] focused on cold start link prediction.

Jiang et al. [43] proposed a user topic based on collaborative filtering approach for personalized travel recommendation which is the improved version of traditional for collaborative filtering by fusing user information in social media.

III. PROPOSED APPROACH

The proposed service quality evaluation is done by three different steps,

- Users' confidence value
- Contextual Features of User Ratings
- Service Quality Evaluation Model

In the users' confidence value information entropy is calculated and in the Contextual user information the features like spatial-temporal features and sentimental features are obtained. Finally in the service quality evaluation model the above two steps are fused together to obtain the overall confidence rating of an item

3.1 Users' Confidence Value

Different users' have different contribution in quality evaluation. In this paper, user rating confidence is leveraged to conduct service quality. Entropy is a measure of uncertainty. The information entropy is used to calculate the confidence value. The difference between the user rating and overall rating reflects the stability of the system. Additionally, the coefficient is added to distinguish the weights because entropy cannot make any difference. If the entropy value is low then the system is more stable so the reciprocal of entropy value is taken.

$$E_u = -\frac{1}{\sum_i(|d_i| \times p(d_i) \log_2 p(d_i))} \tag{1}$$

$$d_i = r_{u,i} - r_i \tag{2}$$

Where E_u denotes user $u's$ confidence value. d_i is the difference between user rating $r_{u,i}$ and overall rating r_i. $p(d_i)$ indicates the probability of the value d_i.
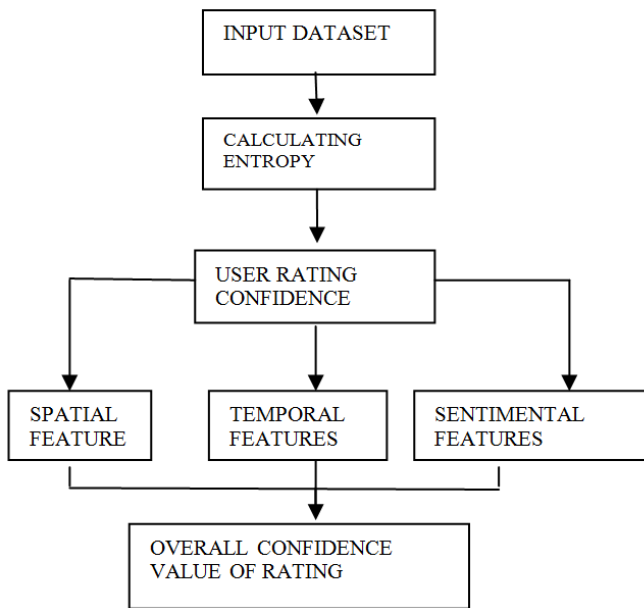
Figure 1. Architecture of Service Quality Evaluation Model

3.2 Contextual Features of User Ratings

The entropy calculation of user ratings confidence is based on the ratings of the user. User profile changes constantly so that their rating's confidence may be different at different places and different time. Sometimes, user gives high rating but there are many negative words in their review. Thus, we further constrain each rating's confidence by its spatial-temporal features and review sentimental features.

3.2.1 Spatial Features

In large network people are living and so they may be influenced by others easily. We start by analyzing the distribution of rating's confidence in different user-item geographic location distances. The user-item geographic distance is calculated by following algorithm:

$$x = lnD(u, i) \tag{3}$$

Where D (u,i) denotes the geographical distance between user u and item i. If the users are close to rated item then the rating's confidence is low. The users may be influenced by their friends or some discounts for services. In terms of items, most of them have competitors.

Generally, competitors are close and mostly native geographically. For different datasets the spatial features are different. Therefore, curve fitting is conducted to learn rating's spatial features. The curve fitting model is based on the 4th Gaussian degree of model. The curve fitting model formula is:

$$y = \sum_j a_j \times exp\left(-\left((x - b_j)/c_j\right)^2\right) \tag{4}$$

Where a_i, b_i, and c_i are the coefficients which is learned in curve fitting. Rating confidence is inversely proportional to y. The rating confidence based on spatial features is represented by:

$$G_{u,i} = 1/\sum_j a_j exp\left(-\left((\ln D(u, i) - b_j)/c_j\right)^2\right) \tag{5}$$

Where $G_{u,i}$ denotes rating confidence user u to item i based on spatial features. a_i, b_i, and c_i are the coefficients which is learned in curve fitting. D(u,i) denotes the geographical distance value between user u and item i.

3.2.2 Temporal Features

In the same way we calculate the rating's confidence based on temporal features. For a single item there are more and more ratings and reviews which result in getting more and more information from former ratings reviews, and then give a suitable rating.

Curve fitting is conducted based on 4th degree Gaussian model. Rating's temporal features can be represented by:

$$T_{u,i} = 1 / \sum_j a_j exp\left(-\left((Day(u,i) - b_j)/c_j\right)^2\right) \quad (6)$$

Where $T_{u,i}$ denotes rating confidence user u to item i based on temporal features. a_i, b_i, and c_i are the coefficients which is learned in curve fitting. $Day(u,i)$ denotes the rating time of user u to item i.

3.2.3 Sentimental features

In most review web sites users not only rate the commodity but also share their experiences and attitude by reviewing. From the textual reviews, we can get exact information, which verifies and supports the rating directly. It is necessary to analyze the relevance between user confidence and textual review sentiment.

First, the method of sentiment analysis is used to calculate sentiment scores. Second, the relevance between user rating confidence and review sentimental is mined. Last, we learn sentimental features to constrain user's confidence. The overall rating of service decrease with the sentiment score. The user confidence increases with review sentiment score. The sentimental features can be represented by:

$$S_{u,i} = 1 / \sum_j a_j \times (RS(u,i))^2 \quad (7)$$

Where $S_{u,i}$ denotes rating's confidence user u to item i according to review sentimental features. $RS(u,i)$ is the normalized sentiment score user u to item i.

3.3 Service Quality Evaluation Model

In this model, we fuse user's confidence with contextual features, including spatial-temporal features and review sentimental features to calculate the overall confidence value of rating. The coefficient is defined in such a way that the sum of coefficient is one. By using probabilistic unified model the spatial-temporal and sentimental features are calculated. The overall confidence of the rating that user u to item i as follows:

$$\phi_{u,i} = A_{u,t(u,i)}T_{t(u,i)} + B_{u,g(u,i)}G_{g(u,i)} + C_{u,s(u,i)}S_{s(u,i)} + D_{u,t(u,i),g(u,i),s(u,i)}E_u \quad (8)$$

Where

$$D_{u,t(u,i),g(u,i),s(u,i)} = 1 - A_{u,t(u,i)} - B_{u,g(u,i)} - C_{u,s(u,i)} \quad (9)$$

Where $t(u,i)$ denotes the time user u rated item i. $g(u,i)$ denotes the geographic distance between user u to item i. $s(u,i)$ denotes the sentimental value of the review. $T_{t(u,i)}$ denotes the temporal value of the review. A, B, C, D are the corresponding coefficients matrices.

3.3.1 Model Inference

The Gaussian with the probabilistic linear model is chosen [23], [25] and [31]. The conditional probability of the observed rating is as follows:

$$p(R|A,B,C,D,G,E,T,S,\sigma_R^2) = \prod_i N\left(R_i \Big| \sum_{u=0}^{n_i}\left(\frac{\phi_{u,i}}{\sum_{u=0}^{n_i}\phi_{u,i}}r_{u,i}\right), \sigma_R^2\right) \quad (10)$$

Where $N(x|\mu,\sigma^2)$ denotes the probability density function with mean μ and variance σ^2. A, B, C, D is user's temporal, spatial, sentimental and coefficients matrix. If there is only one user having rated item i, quality evaluation of the service cannot be performed.

According to [31], zero mean Gaussian priors are assumed for user's spatial-temporal and sentimental coefficients vectors:

$$p(A|\sigma_A^2) = \prod_u N(A_u|0, \sigma_A^2) \quad (11)$$

$$p(B|\sigma_B^2) = \prod_u N(B_u|0, \sigma_B^2) \quad (12)$$

$$p(C|\sigma_C^2) = \prod_u N(C_u|0, \sigma_C^2) \quad (13)$$

3.3.2 Model Training

The gradient, we update the coefficient matrices as follows:

$$A_{u,t(u,i)} = A_{u,t(u,i)} - \alpha \frac{\partial \Psi}{\partial A_{u,t(u,i)}} \quad (14)$$

$$B_{u,g(u,i)} = B_{u,g(u,i)} - \alpha \frac{\partial \Psi}{\partial B_{u,g(u,i)}} \quad (15)$$

$$C_{u,s(u,i)} = C_{u,s(u,i)} - \alpha \frac{\partial \Psi}{\partial C_{u,s(u,i)}} \quad (16)$$

Where α is the learning rate.

The service quality evaluation is conducted by the following coefficient matrix as follows:

$$\hat{r}_i = \sum_{u=0}^{n_i} \frac{\phi_{u,i}}{\sum_{u=0}^{n_i} \phi_{u,i}} r_{u,i} \qquad (17)$$

Algorithm 1.Service Quality Evaluation (SQE) Model

Input: Rating matrix **R** in training dataset

      User confidence **E** calculated by equation (1)

      Spatial bias G calculated by equation (5)

      Temporal bias T calculated by equation (6)

      Sentimental bias S calculated by equation (7)

Output: Quality evaluation of test services.

1: Calculate entropy using difference between user rating and overall rating.

2: Compute the distance for user u and item i

3: Evaluating the temporal features by using curve fitting

4: Then performing sentimental review using normalized sentiment score

5: Fuse all the steps from (1)-(4) into a model to calculate overall rating

6: Initialize coefficients matrices A, B, C, set learning rate α

7: **for** t=1:T **do**

8: **for** each element of coefficients matrices A, B, C, **do:**

9: Using equation (14), (15), (16) the coefficient matrices are updated

10: **end for**

11: **end for**

12: **for** each test item **do**

13: **for** each rating of this item **do**

    Calculate the overall confidence by Equation (8);

14: **end for**

15: calculating the overall rating of the item

16: **end for**

17: **Return**: The overall rating of services

IV. DATASET DESCRIPTIONS

Yelp dataset and Douban dataset is introduced in this section. The dataset can be downloaded from web site of SMILES LAB[1].

4.1 Yelp Dataset

Yelp is a local directory service with social networks and user reviews. It is the largest review site in America.

Users rate the businesses, submit comments, communicate, experience, etc. It combines local reviews and social networking functionality to create a local online community. In our work, we utilize two categories: Restaurant and Nightlife. Moreover, it is proved by the data of Yelp that users are more willing to visit places or to consume items that his/her friends have visited or consumed before. We analyze the relevance between user ratings and user-item location distances.

4.2 Douban Dataset

Douban is one of the most popular social networks in china. It includes several parts: Douban Movie, Douban Read and Douban Music, etc. We crawled the ratings from the Douban Movie websites. The dataset consists of 2,968,648 ratings from 8,226 users who have rated 14,715 movies. Note that there is no geographic location information and reviews in Douban dataset. We perform our model on Douban dataset by fusing user ratings' confidence and temporal features.

4.3 Pre-processing

The issue proposed in this paper is quality evaluation for services with very few ratings. The ratings in our dataset are split according to preselected items. Every tested item will not have more than five ratings. Some items are used for training and some item are used for testing.

V. EXPERIMENTS

The predicted overall ratings of services, the performance of methods will be embodied by the errors. The differences between the prediction and the overall rating of services can be leveraged to measure the model. The real overall ratings of services are discrete as [1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0], while the predictions are in decimals. The predicted decimals can be rounded into discrete quantities. Then precision, Recall and AUC (Area under Curve) measures [47], [48], [49] are utilized to evaluate the proposed model. The proposed service quality is implemented in NetBeans using JAVA language.

VI. CONCLUSIONS

Many researches are focused on rating prediction and personalized recommendation. So, it is important to conduct service quality evaluation. In this paper, we propose service quality evaluation by exploring user's contextual information. We focused on exploring user rating's confidence. The spatial-temporal and sentimental features are calculated. Finally all are fused together to calculate the overall rating confidence. We

use a few ratings to predict the overall ratings of the services.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005.

[2] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in Proc. 10th Int. Conf. World Wide Web, 2001, pp. 285–295.

[3] M. Jahrer, A. Toscher, and R. Legenstein, "Combining predictions for accurate recommender systems," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 693–702.

[4] T. Ma, et al., "Social network and tag sources based augmenting collaborative recommender system," IEICE Trans. Inf. Syst., vol. E98 D, no. 4, pp. 902–910, 2015.

[5] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," J. Mach. Learn. Res., vol. 11, pp. 2057–2078, 2010.

[6] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2008, pp. 426–434.

[7] Y. Koren, "Collaborative filtering with temporal dynamics," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2009, pp. 447–456.

[8] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating collaborative filtering recommender systems," ACM Trans. Inf. Syst., vol. 22, no. 1, pp. 5–53, 2004.

[9] N. Liu, M. Zhao, and Q. Yang, "Probabilistic latent preference analysis for collaborative filtering," in Proc. 18th ACM Conf. Inf. Knowl. Manage., 2009, pp. 759–766.

[10] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 9, pp. 2546–2559, Sep. 2016.

[11] Y. Chen and J. Canny, "Recommending ephemeral items at web scale," in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2011, pp. 1013–1022.

[12] M. Harvey, M. Carman, I. Ruthven, and F. Crestani, "Bayesian latent variable models for collaborative item rating prediction," in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage., 2011, pp. 699–708.

[13] X. Yang, Y. Guo, and Y. Liu, "Bayesian-inference based recommendation in online social networks," in Proc. IEEE INFOCOM, 2011, pp. 551–555.

[14] H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King, "Recommender systems with social regularization," in Proc. 4th ACM Int. Conf. Web Search Data Mining, 2011, pp. 287–296.

[15] L. Yu, R. Pan, and Z Li, "Adaptive social similarities for recommender systems," in Proc. 5th ACM Conf. Recommender Syst., 2011, pp. 257–260.

[16] P. Bedi, H. Kaur, and S. Marwaha, "Trust based recommender system for Semantic Web," in Proc. 20th Int. Joint Conf. Artifical Intell., 2007, pp. 2677–2682.

[17] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun, "Who should share what? Item-level social influence prediction for users and posts ranking," in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2011, pp. 185–194.

[18] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting place features in link prediction on location-based social networks," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 1046–1054.

[19] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint social and content recommendation for user-generated videos in online social network," IEEE Trans. Multimedia, vol. 15, no. 3, pp. 698–709, Apr. 2013.

[20] Y. Chen, A. Cheng, and W. H. Hsu, "Travel recommendation by mining people attributes and travel group types from community- contributed photos," IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1283–1295, Oct. 2013.

[21] G. Zhao, X. Qian, and X. Xie, "User-service rating prediction by exploring social users' rating behaviors," IEEE Trans. Multimedia, vol. 18, no. 3, pp. 496–506, Mar. 2016.

[22] G. Zhao, X. Qian, and C. Kang, "Service rating prediction by exploring social mobile users' geographic locations," IEEE Trans. Big Data, to

be published. Doi: 10.1109/TBDATA.2016.2552541.

[23] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 1267–1275.

[24] M. Jiang, et al., "Social contextual recommendation," in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage., 2012, pp. 45–54.

[25] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," IEEE Trans. Knowl. Data Eng., vol. 26, no. 7, pp. 1487–1502, Jul. 2014.

[26] H. Feng and X. Qian, "Recommendation via user's personality and social contextual," in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., 2013, pp. 1521–1524.

[27] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews," IEEE Trans. Multimedia, vol. 18, no. 9, pp. 1910–1921, Sep. 2016.

[28] G. Zhao and X. Qian, "Service objective evaluation via exploring social users' rating behaviors," in Proc. IEEE Int. Conf. Multimedia Big Data, 2015, pp. 228–235.

[29] Y. Koren, "Collaborative filtering with temporal dynamics," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2009, pp. 447–456.

[30] G.Dror, N. Koenigstein, and Y. Koren, "Yahoo! Music recommendations: Modeling music ratings with temporal dynamics and item taxonomy," in Proc. 5th ACM Conf. Recommender Syst., 2011, pp. 165–172.

[31] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in Proc. 4th ACM Conf. Recommender Syst., 2010, pp. 135–142.

[32] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in Proc. NIPS, 2007, pp. 1257–1264.

[33] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[34] K. Lee and K. Lee, "Using dynamically promoted experts for music recommendation," IEEE Trans.

Multimedia, vol. 16, no. 5, pp. 1201–1210, Aug. 2014.

[35] X. Yang, T. Zhang, and C. Xu, "Cross-domain feature learning in multimedia," IEEE Trans. Multimedia, vol. 17, no. 1, pp. 64–78, Jan. 2015.

[36] X. Wang, et al., "Semantic-based location recommendation with multimodal venue semantics," IEEE Trans. Multimedia, vol. 17, no. 3, pp. 409–419, Mar. 2015.

[37] L. Hu, A. Sun, and Y. Liu, "Your neighbors affect your ratings: On geographical neighborhood influence to rating prediction," in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 345–354.

[38] W. Zhang, G. Ding, L. Chen, C. Li, and C. Zhang, "Generating virtual ratings from chinese reviews to augment online recommendations," ACM Trans. Intell. Syst. Technol., vol. 4, no. 1, 2013, Art. no. 9.

[39] S. Tan, et al., "Interpreting the public sentiment variations on Twitter," IEEE Trans. Knowl. Data Eng., vol. 26, no. 5, pp. 1158–1170, May 2014.

[40] S. Chelaru, I. Altingovde, S. Siersdorfer, and W. Nejdl, "Analyzing, detecting, and exploiting sentiment in web queries," ACM Trans. Web, vol. 8, no. 1, 2013, Art. no. 6.

[41] V. Leroy, B. Cambazoglu, and F. Bonchi, "Cold start link prediction," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 393–402.

[42] P. Lou, G. Zhao, X. Qian, H. Wang, and X. Hou, "Schedule a rich sentimental travel via sentimental POI mining and recommendation," in Proc. IEEE 2nd Int. Conf. Multimedia Big Data, 2016, pp. 33–40.

[43] D. Quercia, N. Lathia, F. Calabrese, G. Lorenzo, and J. Crowcroft, "Recommending social events from mobile phone location data," in Proc. 2010 IEEE Int. Conf. Data Mining, 2010, pp. 971–976.

Multi-focus Image Fusion with Quantitative Analysis

S. Abirami, G. Rajasekaran

Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi, TamilNadu , India

ABSTRACT

Multi-focus image fusion is the combining relevant information from two or more images of a same scene and as results has "all-in-focus" image. When one scene contains objects in different distance, the camera can be focused on one another of each object, generate set of pictures. Then, by applying image fusion techniques, an image with better focus across all area can be generated. This paper describes an image fusion system using different fusion techniques and the resultant is analyzed with quantitative measures. Initially, the registered images from two different modalities are considered as input image. For the resultant data the perceptual image fusion is applied and the fused image is analyzed with quantitative metrics namely Peak Signal –to- Noise Ratio (PSNR), Mutual Information (MI), Structural Similarity Index (SSIM). From this experimental result we observed that the proposed fusion method provides better result compared to the given images as justified by quantitative metrics.

Keywords: Image Fusion; Discrete Wavelet Transform; Dual Tree Complex Wavelet Transform; Quantitative Metrics

I. INTRODUCTION

The effective fusion for more than one resource with high resolution and having high information content in context with visualization, scene understanding, target recognition and situational awareness in multi-sensor applications such as medicine, surveillance and remote sensing. In the research of fusion techniques, there is several fusion techniques are proposed and implemented, but these techniques have certain limitations. Some of the image fusion techniques are available in Fusion Tool of Matlab5.0 namely are Filter-subtraction-Decimate Pyramid (FSD), Gradient Pyramid, Laplacian Pyramid [4], Discrete Wavelet Transform Pyramid (DWT), Shift Invariant Discrete Wavelet Transform Pyramid (SIDWT) [3], Principle Component Analysis [2], Morphological Pyramid, Contrast Pyramid, Ratio Pyramid, and so on [1]. For example, Contrast Pyramid loses too much information from the source images; Lots of false information produced by the Ratio Pyramid method which never existed in the original images; Many false edges created by the Morphological Pyramid method. In a word, these methods cannot deal with various types of images. Many of the fusion techniques are based on wavelet transformation. But, the DWT image fusion method is resulting with shift variant and additive noise in fused image. Using Redundancy Discrete Wavelet Transform (RDWT), Contourlet Transform [5] and Dual-Tree Complex Wavelet Transform (DTCWT) [6]. An RDWT fusion method is used to preserve the exact edge and spectral information from the given images without any loss of spatial information. In this technique, the high pass and low pass sub bands of the input images are fused using the average method and entropy method respectively. The region based Contourlet Transform gives local brightness, localization, multiresolution, directionality and anisotropy, etc. on the fused image. This transformation process is implemented in two stages: a) transformation and b) subband decomposition. On the first stage, double filter bank is applied and the second stage, local energy is calculated to the each subband and then fusion rules are applied like average mode and selection mode. DTCWT has good directional selectivity as compare to other methods and it also reduced shift variant property.

Following are features of DTCWT:

- Good directional selectivity
- Approximate shift variant
- Perfect reconstruction using short linear filters
- Limited redundancy
- Efficient order n computations

The actual fusion process should retain as much perceptually important information as possible from the two sources and should form a single more informative image [1]. In the transform domain, firstly the input images are decomposed based on transform coefficients. Then the fusion technique is applied to obtain the fusion decision map. The inverse transformation on this decision map yields the fused image. The fused image contains all the information of the source images and reduces the spatial distortion.

The performances of different image fusion methods are analysed using different measures like quantitative metrics namely PSNR, SSIM and MI.

The rest of the parts are follows: In section II, the logical design is briefly reviewed; experimental results are described in section III; the performance of the proposed method is evaluated based on the quantitative metrics; section IV contains the conclusion.

II. LOGICAL DESIGN

In this system initially the two images are taken as input. Then the fusion techniques are applied to the registered images to find a more informative fused image. Finally, fused image information is validated using the quantitative measures. The overall system design is shown in Fig.1.

A. Dataset

The fusion image used by petrovic [11] is taken as the dataset (comprising remote sensing, medical and visual/thermal pairs). Each set of given images were taken from the same scene but at the different views.
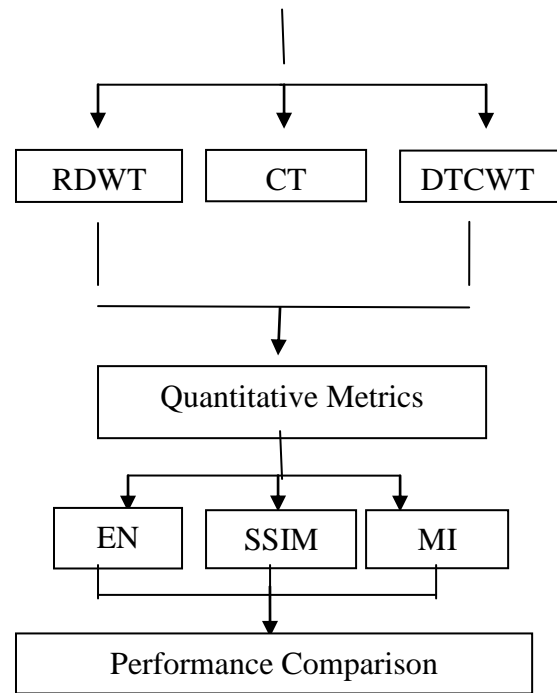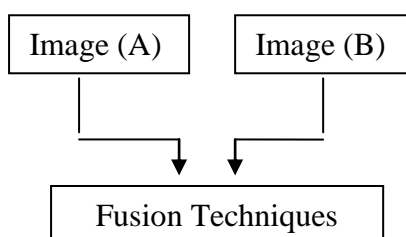
Figure 1. An overall view of logical design

B. Redundancy Discrete Wavelet Transform

RDWT fusion technique can be effectively performed on two registered images with captured at different time instance.

Let A and B be the two registered images of two different modalities. The registered images are decomposed into three levels of RDWT decomposition using Daubechies filters on the both input images in order to produce an approximate wavelet bands as shown in Fig. 2
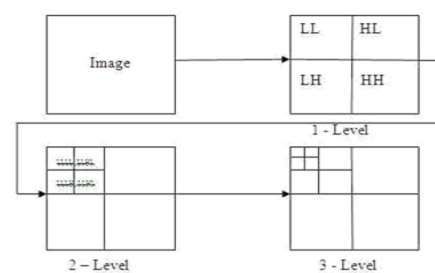
Figure 2. Three levels of decomposition

In the first phase I1 image is decomposed into $I_A^p, I_A^q, I_A^r, I_A^s$ be the RDWT subband and also $I_B^p, I_B^q, I_B^r, I_B^s$ be the corresponding RDWT subbands from I2 image. To extract the features from both the images, coefficients from approximation band of IA and IB are averaged;

$$I_N^a = mean(I_A^p, I_B^q) \qquad (1)$$

Where I_N^a is the approximation band of the fused image.

In the next phase each subband namely LH, HL, HH, HL is divided into blocks of size 3*3 and the entropy of each block is calculated, as in (2)

$$e_i^{ab} = \ln \sqrt{(\mu_i^{ab} - \Sigma_{x,y=1}^{3,3} I_i^{ab}(x,y)/\sigma_i^{ab})^2 / r^2} \qquad (2)$$

Where (a = q, r, s) specify the subbands, m = 3 (size of each block), b represents the block number, and (i = A, B) is used to differentiate the two multimodal images I_A and I_B. μ_i^{ab} and σ_i^{ab} are the mean and standard deviation of the RDWT coefficients. Using the entropy values, the detail subbands for the fused image I_N^q, I_N^r, I_N^s, and I_N^h are generated, the derived fused image block I_N^{ab}, RDWT coefficients from I_A image is greater than the specific block of I_B image, otherwise I_B^{ab} is selected.

$$I_N^{ab} = \begin{cases} I_A^{ab}, & \text{if } (e_1^{ab} > e_2^{ab}) \\ I_B^{ab}, & otherwise \end{cases} \qquad (3)$$

Finally, IRDWT is applied on all the subbands to generate the resultant fused image I_N.

$$I_N = IRDWT(I_N^p, I_N^q, I_N^r, I_N^s) \qquad (4)$$

C. Contourlet Transform

The contourlet transform is a two-dimensional transform method for image representations. The CT has properties of multiresolution, directionality, localization, critical sampling and anisotropy. Its basic functions are multidimensional and multiscale.

Contourlet transform causes smoothness in a fused image with any two different modalities of images [6]. This transformation process involved in two stages. Double filter bank is applied for transformation in the first stage and decomposition process is done with fusion rules in the second stage. Finally, the fused output image is efficiently retrieved by using reconstruction procedure.

The block diagram of the proposed image fusion algorithm is shown in Fig. 3. Here images A and B

represent the input source image. F is the final outcome of the fused image after applying the inverse contourlet transform.
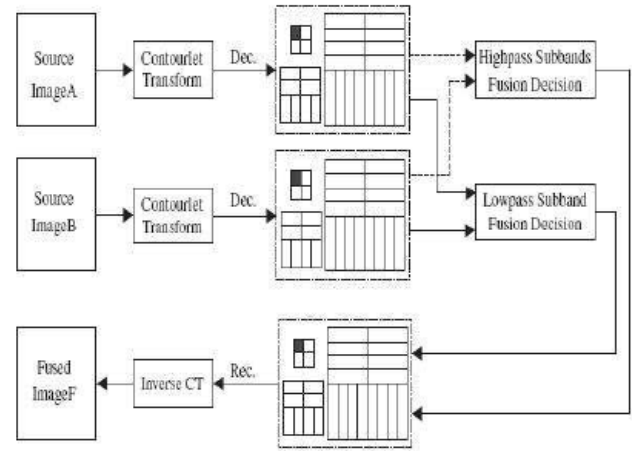
Figure 3. Flow diagram of the contourlet based image fusion

1) Transformation Stage

A double filter bank scheme is utilized efficiently for subband decomposition using Laplacian Pyramid (LP) and Directional Filter Bank (DFB) in this stage. To capture the edge point Laplacian Pyramid is used. To link the discontinuities point in linear structures Directional Filter Bank is used.

Each input image is decomposed into a subband of low frequency of original image and a bandpass high frequency subbands in Laplacian Pyramid method [7]. The same process is repeated for the specified contourlet decomposition level for low frequency subband. The decomposition process of Laplacian Pyramid is shown in Fig. 4. First the input image is applied to a LP filter H and then down sampled to derive a coarse approximation a. After performing the down sampled the image is up sampled and passed through a synthesis filter G. The result of highpass subbands are derived from subtracting the output of the synthesis filter with the input image.
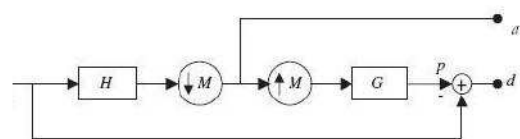
Figure 4. Construction of LP

2) Decomposition Stage

The decomposed subbands of transformation stage are merged using lowpass and highpass fusion rule.

a) Lowpass subband fusion:

The coefficients in the coarest scale subband a represents the approximation component of the source image. In this method, the local energy contourlet domain is developed as the measurement, then the selection mode and averaging modes are used to compute the final coefficients [7].

Figure 5. Construction of DFB

The local energy E(x,y) is calculated the current Coefficient in the approximate subband a, which is

$$E(x,y) = \Sigma_m \Sigma_m a_j (x+m, y+n)^2 W_L(m,n) \quad (5)$$

Where (x,y) is the current contourlet coefficient, $w_L(m,n)$ is a template of size 3*3

$$W_L = \frac{1}{9} * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (6)$$

The salience factor is calculated to identify whether the selection mode and averaging mode to be used in the fusion process.

$$M_j^{AB}(x,y) = 2\sum_m\sum_n a_j^A(x+m,y+n)a_j^B(x+m,y+n)/E^A(x,y)+E^B(x,y) \quad (7)$$

Where $a_j^x(x,y)$; x=A, B is the lowpass contourlet coefficients of the source image A or B and $M_j^{AB}(x,y)$ is the salience factor.

Similarity of the lowpass subbands of the two source images are reflects by the salience factor. After that, this value is compared to a predefined threshold T_L

$(T_L = 1)$.The averaging mode is chosen for the condition $M_j^{AB}(x, y) > T_L$ with the fusion rules denoted in (8).

$$a_j^F(x, y) = \alpha_A.a_j^A(x, y) + \alpha_B.a_j^B \quad (8)$$

Where $a_j^F(x, y)$ denotes the fused result at position (x,y). The α_A and α_B are selected based on the specified condition, as in (9).

$$\alpha_A = \alpha_{min} \text{ for } E^A(x, y) < E^B(x, y)$$
$$= \alpha_{max} \text{ for } E^A(x, y) \geq E^B(x, y) \quad (9)$$

Where $\alpha_B = 1 - \alpha_A, \alpha_{min} \in (0,1)\alpha_{min} + \alpha_{max} = 1$

If $M_j^{AB}(x, y) \leq T_L$, selection mode is selected for fusion in (10).

$$a_j^F(x, y) = a_j^A(x, y) \text{ for } E^A(x, y) \geq E^B(x, y)$$
$$= a_j^B(x, y) \text{ for } E^A(x, y) < E^B(x, y) \quad (10)$$

b) Highpass subband fusion:

The coefficients which have larger absolute values in the high frequency subbands $d_{j,k}$ are fused using the average method is defined as follows

$$E_{j,k}^F(x, y) = d_{j,k}^A(x, y) + d_{j,k}^B(x, y) \quad (11)$$

Where $E_{j,k}^F(x, y)$ denotes the local energy, $d_{j,k}^X(x, y)$ is the high frequency coefficient.

c) Reconstruction of fusion image:

The fused image is obtained from $a_j^F(x, y)$ and $E_{j,k}^F(x, y)$ using inverse contourlet decomposition method.

D. Dual-Tree Complex Wavelet Transform

To overcome the drawbacks of DWT, in 1998 kingsbury proposed the dual-tree complex wavelet transform (DTCWT), which provides both good shift variance and directional selectivity. The DTCWT design is based on the use of two parallel trees, first one for odd samples and the second one for the even samples generated at the first stage. These trees provide the signal delays necessary for every level and consequence eliminate aliasing effect. The subbands are divided into six distinct

subbands which are $\pm15, \pm45, \pm75$.

A. Perceptual Fusion Framework for DTCWT

The calculation of JND model proposed by Liu et al. [8] where the local JND associated with a coefficient is defined as:

$$t_{JND}(\lambda, \theta, r, s) = JND_{\lambda, \theta} a_l(\lambda, \theta, r, s) a_c(\lambda, \theta, r, s) \quad (12)$$

Where $JND_{\lambda, \theta}$ the base detection threshold is for DTCWT subband at level λ and orientation θ; $a_c(\lambda, \theta, r, s)$ is the luminance masking and $a_c(\lambda, \theta, r, s)$ is the contrast masking.

a) Luminance Masking:

The contrast of human visual system is dependent on the luminance context. These model proposed by Chou and Li [9].

$$a_l = \begin{cases} 17\left(1 - \sqrt{\dfrac{\overline{I}}{127}}\right) + 3, & \overline{I \le 127} \\ \dfrac{3}{128}\left(\overline{I} - 127\right) + 3, & otherwise \end{cases} \quad if \quad (13)$$

$\overline{I}$ is calculated from the magnitude of the co-located lowpass coefficient at the highest decomposition level.

a) Contrast Masking:

The contrast masking modelled a_c can be written as:

$$a_c(\lambda, \theta, r, s) = a_{c\_intra}(\lambda, \theta, r, s) a_{c\_inter}(\lambda, \theta, r, s) \quad (14)$$

The intra band contrast masking can be modelled by the nonlinear transducer model introduced by Teo [10], which is written as:

$$a_{c\_intra}(\lambda, \theta, r, s)$$
$$= \max\left\{1, W_{intra} \sum_{v \varepsilon C_{i,j}(h)} \left|\frac{v}{JND_{\lambda, \theta}}\right|^{\varsigma} \Big/ N_{r,s}\right\}, \quad (15)$$

Where (λ, θ, r, s) is the DTCWT coefficient (λ, θ, r, s). The weighting factor W_{intra} and exponent factor ς is set to 12 and 0.6.

The inter band contrast masking written as:

$$a_{c\_inter}(\lambda, \theta, r, s)$$
$$= \max\left\{1, W_{inter} \sum_{v \in C\lambda, \theta} \omega_\lambda \omega_\theta \left|\frac{v}{JND_{\lambda, \theta}}\right|^{\beta} \Big/ N_{r,s}\right\} \quad (16)$$

Where (λ, θ, r, s) is the DTCWT coefficient (λ, θ, r, s). The weighting factor W_{inter} and exponent factor ς is set to 12 and 0.6.

E. Quantitative Analysis

For the resultant data we are checking the quality of fusion technique by comparing the with various methods by the quantitative measurement. The few metrics explains the quantitative metrics in the following section.

1) **Peak Signal to Noise Ratio (PSNR):** To measure the quality of the image with respect to the original input image PSNR is used.

$$MSE = \frac{1}{pq} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} [M(i,j) - N(i,j)]^2 \quad (17)$$
$$PSNR = 10 \log_{10}(MAX^2 / MSE) \quad (18)$$

Where MAX denote the maximum value in an image. p, q are the height and weight of an image. M(i,j) is the input image and N(i,j) is the fused image.

2) **Structural Similarity Index (SSIM):** SSIM is used to measure the similarity between two images. It is defined as

$$SSIM(A, B) = (2(\mu_A \mu_B + E_1) * 2(\sigma_{AB} + E_2)) /$$
$$((\mu_M^2 + \mu_N^2 + E_1) * (\sigma_M^2 + \sigma_N^2 + E_2)) \quad (19)$$

Where μ_M and μ_N is the mean intensities, σ_M and σ_N is the standard deviation, σ_{MN} gives the covariance of M and N, E_1 and E_2 are constants.

3) **Mutual Information (MI):** let A and B be the two registered images, then the mutual information is given as

$$MI(M, N) = E(M) + E(N) - E(M, N) \quad (20)$$

Where E(M) is the entropy of image M, E(N) is entropy of image N and E(M,N) gives the joint entropy.

III. EXPERIMENTAL ANALYSIS

The following section describes the sequence of the fusion of input images with their performance analysis.

STEP 1: Two given images are taken as test images. Each image have the same pixel size such as 256*256, with 256-level gray scale. Totally, five set of images are used in analyze the performance. One set of input images shown in fig. 6
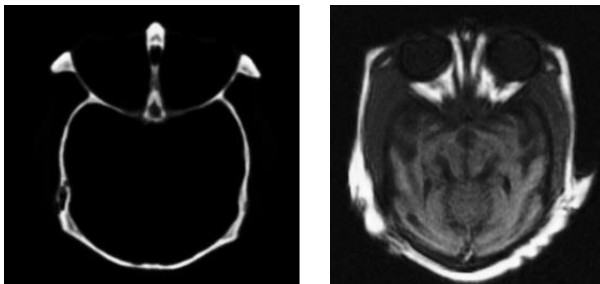
Figure 6. Dataset 1

STEP 2: The input images of each dataset are fused using RTDWT, CT and DTCWT and some of the samples are shown in Fig.7
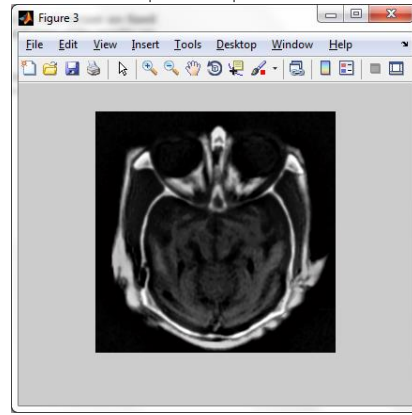
Figure 7. Fused Image of Dataset 1 using DTCWT

STEP 3: Finally, the performance of the fused image is obtained from each technique are analysed with subjective metrics discussed in section II. The evaluation performance of all the techniques, analyzed for the output image with the quantitative metrics are shown in the Table 1

METRICS	ALGORITHM	INPUT SET 1	INPUT SET 2	INPUT SET 3
EN	RDWT 1	6.6693	6.6571	6.5767
	RDWT 2	6.7634	6.7412	6.6638
	RDWT 3	6.8405	6.7866	6.7193
	CT	6.3936	6.1643	6.3261
	DTCWT			
SSIM	RDWT 1	54.6854	53.0163	54.8240
	RDWT 2	59.4352	58.6768	60.1467
	RDWT 3	60.7343	59.0563	60.0563
	CT	56.6345	55.6450	56.6890
	DTCWT	64.7320	66.7854	70.6995
MI	RDWT 1	60.7703	60.0502	60.9051
	RDWT 2	60.4853	59.8634	60.5467
	RDWT 3	60.1785	59.2891	60.2885
	CT	59.1723	59.3468	59.4563
	DTCWT	62.9325	63.0753	62.2335

IV. CONCLUSION

In this paper, by using petrovic dataset we have analyzed the three different fusion techniques using quantitative metrics.

The proposed work has addressed by producing a principle model of the perceptual importance of coefficients within image fusion and then evaluating the performance quantitatively and qualitatively across a representative dataset.

V. REFERENCES

[1]. Firooz Sadjadi, "Comparative Image Fusion Analysis", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 3, June 2005

[2]. Huaixin Chen, "A Multiresolution Image Fusion Based on Principle Component Analysis", Fourth International Conference on Image and Grapics, pp. 737-741, August 2007.

[3]. Oliver Rockinger, "Image Sequence Fusion Using a Shift-Invariant Wavelet Transform", International Conference on Image Processing, Vol. 3, pp.288, October 1997.

[4]. Peter J.Burt, Edward H.Adelson, "The Laplacian Pyramid ass a Compacct Image Code", IEEE Transcations on Communications, Vol.31, pp. 532-540, April 1983.

[5]. James E.Fowler, "The Redundant Discrete Wavelet Transform and Additive Noise", IEEE Signal Processing Letters, Vol.12, No.9, September 2005

[6]. Souparnika Jadhav, "Image Fusion Based on Wavelet Transform", International Journal of Engineering Reasearch, Vol. 3, pp. 442-445, July 2014.

[7]. L.yang, B.L.Guo, W.Ni, "Multimodality Medical Image Fusion Based on Multiscale Geometric Analysis of Contourlet Transform", Elsevier Science Publishers, Vol. 72, pp. 203-211, December 2008.

[8]. Z.liu, L.J.Karam, and A.B.Watson, "JPEG2000 encodind with perceptual distortion control," IEEE Transactions on Image Processing, Vol. 15, no.7, pp. 1763-1778, July 2006.

[9]. C.-H. Chou and Y.-C. Li, "A Perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, n0. 6, pp.467-476, 1995.

[10]. P.C.Teo and D.J.Heeger, "Perceptual image distortion", in IEEE International conference on image processing, 1994, pp. 982-986.

[11]. V.Petrovic, "Subjective tests for image fusion evaluation and objective metric validation", Information Fusion, Vol. 8, no. 2, pp. 208-216, 2007

Data Partitioning Method for Mining Frequent Itemset Using MapReduce

R. Divya Bharathi[1], A. S. Karthik Kannan[2], E. Jai Vinitha[3]

[1]*Department of IT, M.Tech, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India*
bharathikrishnan2.db@gmail.com[1]
[2]*Department of IT, Assistant Professor, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India*
karthikkannan@mepcoeng.ac.in[2]
[3]*Department of IT, M.Tech, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India*
evinithait12@gmail.com[3]

ABSTRACT

Existing parallel mining algorithm lacks in communication and mining overhead. To overcome this problem a data partitioning method using MapReduce model is proposed. In this model, three MapReduce tasks are implemented to improve the performance of frequent itemset mining in parallel. In second MapReduce job the mapper perform LSH based approach that integrates the item grouping and partitioning process. The reducer performs FP-Growth based on the partition data to generate all frequent patterns in the data. The main idea of data partitioning is to group relevant transactions and reduce the number of the relevant transaction. Extensive experiments using IBM Quest Market Basket Synthetic Datasets to show that data partitioning is efficient, robust and scalable on Hadoop.

Keywords : Frequent Itemset Mining, Mapreduce Model, Parallel Mining, Data Partitioning.

I. INTRODUCTION

Frequent mining is an important problem in sequence mining and association rule mining. Increasing the speed of FIM is critical because Frequent itemset mining consumes more amount of mining time for its high computation in input and output process. When data in mining applications become very large, sequential frequent mining algorithm suffer from performance when it runs on single node [1] [2]. To overcome the problem of performance distortion therefore a framework using MapReduce a widely adopted programming model for processing big datasets by exploiting the parallelism among computing nodes of a cluster. We describe how to distribute a large dataset over the cluster to balance load across the Hadoop nodes to optimize the performance of parallel FIM.

The frequent itemset mining algorithm can be divided in to two Apriori and FP growth. Apriori is a well known method for mining frequent itemsets in a transactional database. The algorithm works within a multiple pass generation and test framework, comprising the joining and pruning phases to reduce the number of candidates before scanning the database for support counting so each processor has to scan a database multiple times and to exchange an excessive number of candidate itemsets with other processors. Therefore Apriori parallel FIM solution suffer potential problems of high I/O and synchronization overhead, which make very difficult to scale up these parallel algorithm.

1.1 Motivations

The main contributions of this paper are given as follows
1) We develop the parallel frequent itemset mining method using MapReduce programming model.
2) partitioning of data in mapreduce play vital role in the performance while processing large datasets.

1.2 MapReduce Framework

MapReduce is a promising parallel and scalable programming model for data intensive applications. A mapreduce program distributes computation as a sequence of parallel operations on dataset of key/value pairs. MapReduce has two phases namely, the map phase and Reduce phase. The map task splits the input in to a N number of fragments, which are distributed evenly to the first tasks across the cluster nodes by node manager for data processing. The input data is split in to small fragments as key-value pair and each map produces key-value pair as intermediate result. The reduce unit take the intermediate result of the map function as key and list of values and output the collection of values. Both tasks can be performed in parallel. The inputs pairs of map and the output pairs of reduce are managed by an underlying distributed file system. It offers automatic data management, transparent fault tolerant processing and highly scalable. The mapreduce can be an efficient platform for frequent itemset mining in large scale dataset.

Hadoop is open source implementation of mapreduce programming model which relies on its own Hadoop Distributed File System (HDFS). HDFS is designed for storing very large files with streaming data access patterns running on commodity hardware. At the heart of HDFS is a single Name Node a master server that manages the file system. The master node consists of task tracker, job tracker, name node and data node. A slave node acts as both a task tracker and data node.

II. RELATED WORK

Many parallel algorithms have been proposed to enhance the performance of the Apriori frequent itemset mining. In Apriori based frequent itemset mining [3] they proposed three algorithms namely single pass counting (SPC), fixed passes combined counting (FPC) and Dynamic pass combined counting (DPC). SPC finds out frequent k itemset at k-th pass of database. FPC finds out (k-1) and (k+1) and up to (k+m) itemset in a map reduce phase. The third algorithm DPC, considers the workloads of nodes and

find out various length of frequent itemsets as possible in a map-reduce phase and also DPC calculates the candidate threshold and prevents the generation of many false positive candidates. It performs well when compared with the SPC and FPC.

X.Lin proposed Mr. Apriori [4] algorithm which runs on a parallel Map/Reduce framework. Prune (C_k+1) function to remove the non frequent itemset from the transaction. Where this function eliminates redundancy execution and c_k cannot be subset of frequent item sets. This algorithm first calculate frequent itemset for each map node as the time complexity with respect to transaction t, Number of transactions n , Number of item in the transactions m. In second task is to calculate frequent item set with an additional item by joining, sorting and eliminating the duplicated items in each map node. Finally similarity can be calculated at the reduce nodes and eliminate the frequencies that do not meet the minimum support.

S. Hong, Z. Huaxuan, C.Shiping, and H.Chunyan [5] proposed improved FP growth algorithm which combines the sub-tree with same patterns which has high support count. Further it combines with mapreduce computing model MR-IFP (mapreduce-improved FP). It uses the depth first method to mine the frequent itemsets and saves a great deal of space. Built cloud platform to implement the IFP based on linked list Therefore it achieves high efficiency and scalability.

M. Liroz-Gistau, R. Akbarinia, D. Agrawal, E.Pacitti, and P. Valduriez [6] state that Map Reduce jobs are executed over distributed system composed of a name node and data nodes. Input is dividing into several splits and assigned to map tasks. MR-Part A partitioning technique is used for automatic partitioning of mapreduce input phase and a locality scheduling is done at the reduce tasks simultaneously so it reduce the amount of data shuffling between map unit and reduce unit .

L. Zhou, Z. Zhong , J. Chang [7] proposed a Balanced parallel FP growth (BPFP) uses two round of

mapreduce to parallelize FP growth. First phase divides entire mining task in to relatively even sub task to improve parallelization. The second phase divides all the load units in to several groups. It eliminates dependency between parallel tasks. In order to balance the load a work load of each mining unit is calculated and then fairly divides this unit into several groups. By using this balance metric BPFP achieves high parallelization.

III. PROPOSED WORK

3.1 Partitioning of Data

Data partitioning is done using voronoi diagram based partitioning [8]. In this technique spaces are divided into number of parts. In prior a group of set which is referred as pivots (seeds) is chosen at preprocessing phase. For each pivot point there is a corresponding part consisting of all data points closer to it .The split regions for particular pivot are called voronoi cells. For a dataset D set of k pivots are selected then all objects in the Dataset D are partitioned into k disjoint sets (denoted as G_1, G_2,..G_k) and each object in D are assigned to the nearest pivot.

3.2 Distance Metric

In order to compute the similarity between the transactions Jaccard similarity is computed. A Jaccard similarity value 1 indicates two transactions are highly similar. The similarity between of two transactions T_1 and T_2 is given as below,

$$J(T_1, T_2) = \frac{|T1 \cap T2|}{|T1 \cup T2|} \qquad (1)$$

$J(T_1, T_2)$ is a value in between 0 to 1,if it is zero then row set is not equal, if it is 1 then the row set are same.

3.3 K-means selection of pivots

Pivot selection is the main process in partitioning the data. K-means strategy is used to select the pivots. It aims to partition m objects into k clusters. Given a set of objects $(y_1, y_2...y_k)$ are partitioned in to clusters

$c_1, c_2, .c_k$ where each data points belong to a cluster for a particular mean. Pivot selection is carried out preprocessing phase.

3.4 Partitioning Strategies

The two partition strategies are MinHash and LSH-Based partitioning in that MinHash is a basic foundation for Locality Sensitive Hashing.

3.4.1 MinHash

MinHash provides solution to determine the similarity between two sets [9]. MinHashing technique is mainly for dimension reduction large sets are replaced by smaller sets called "signatures". There are two phases to generate signatures.

1) Characteristic matrix
2) MinHash Signatures

Characteristic matrix

Characteristic matrix will be obtained from FList (frequent one itemset List) and original dataset. Where column represent transaction and row denotes items in the transaction. For a given Dataset D= {T_1, T_2, .T_k}, which contains m items. If the item in the FList present in the transaction T_1 then the item alone will be set as 1 otherwise it will be set as 0. Likewise the entire transaction in Dataset will be checked with FList and m (number of items) by n
(number of rows) characteristic matrix of M will be obtained.

3.4.2 MinHash Signatures

Signature matrix will be constructed using characteristic matrix M for every item in the M generates hash function [10]. Initially the signature matrix value will be set as infinity after that if the transaction contains 1 then replace hash value in the signatures. Likewise for each transaction compute signature value. The characteristic matrix and signature matrix would consists of the same number of

columns but very less rows, thereby drastically reducing the dimensions.

LSH-Based partitioning

Based on the banding technique, the rows in the signature matrix are divided into (b x r) where b represents the number of bands. Each band consists of maximum r rows. For each band a hash function is defined. The function takes column of its corresponding band and hashes them to large number of buckets. We can use the same hash function. For a any hash family if any two sets m_1 and m_2 satisfy the below conditions, then hash family is called (R, q, m_1, m_2).

A) If $||m_1 - m_2|| < R$
B) If $||m_1 - m_2|| < cR$

Above condition is used to check similar sets are mapped in to the same buckets and dissimilar sets will be mapped in the individual bucket in hash table

IV. IMPLEMENTATION

Data partitioning phase consist of three phases. In preprocessing phase selection of k pivots will be done at the master node , which will be the input for second phase in mapreduce.

A. First MapReduce job

The first MapReduce phases discovers all frequent one-itemsets. The transaction in the dataset is partitioned into Data Nodes and items in the transaction and items in the transaction are counted parallel.

System Architecture
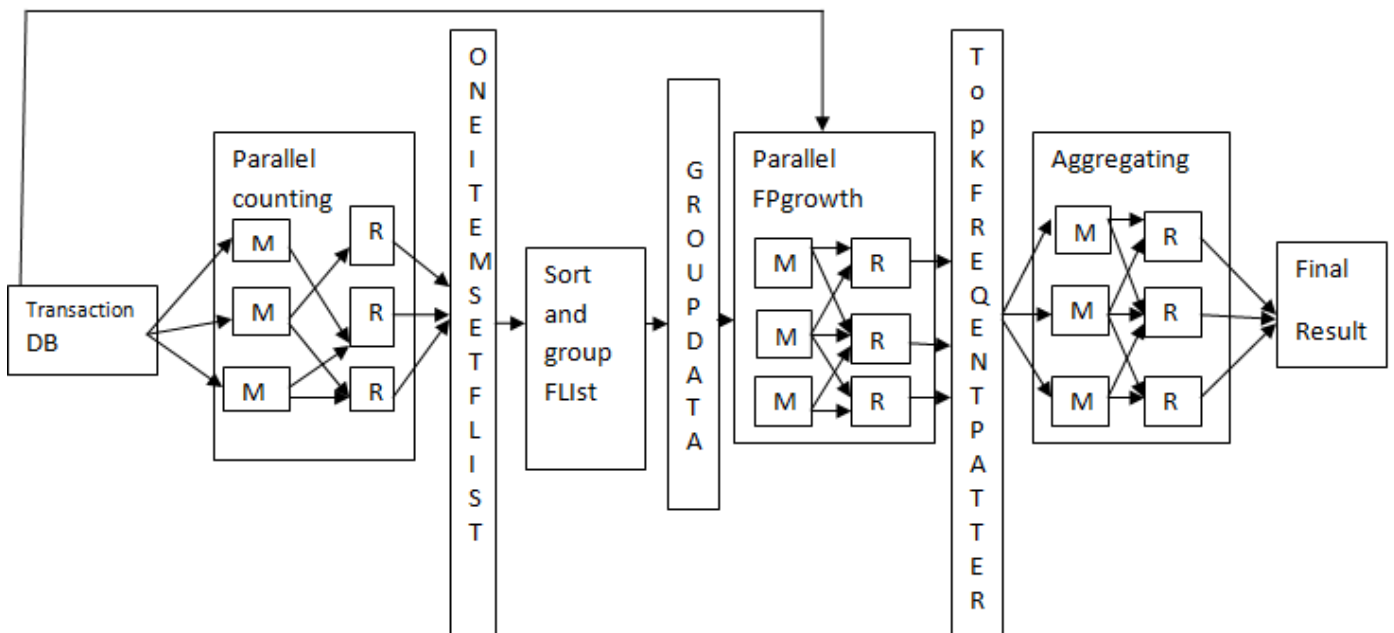
Figure 1. system architecture of pfp
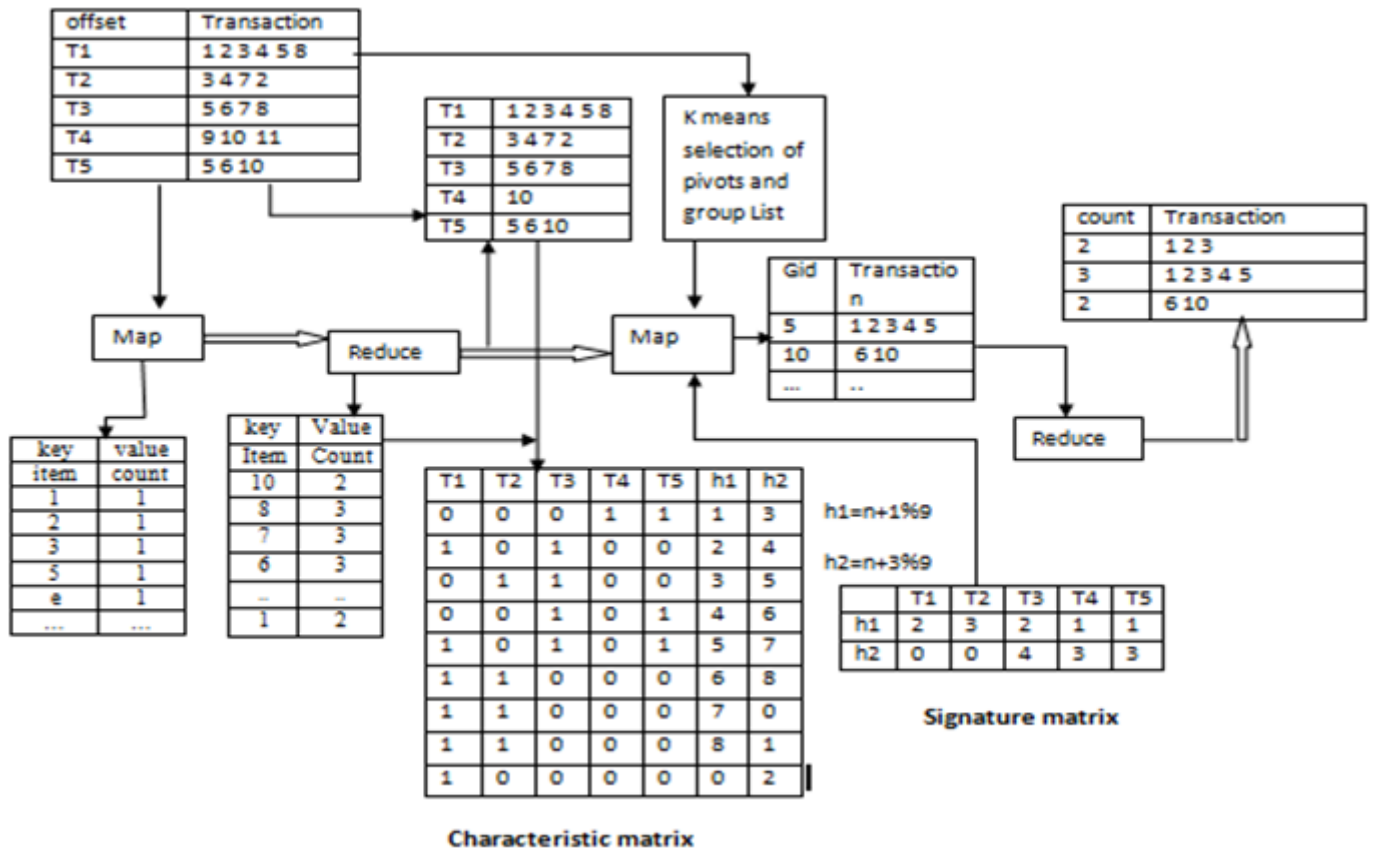
Figure 2. Overview of Data Partitioning Algorithm

The input of Map tasks is a database, and the output of Reduce tasks is one frequent itemsets. The map input format will be <Long Writable offset, Text value> and the reduce output format will be in the format of <Text item, Long Writable count> then the frequent one itemset count along with itemsets are stored in a file called F-List.

B. Second MapReduce job

Mapper

First MapReduce output will be given as an input for the second phase in MapReduce. In second MapReduce phase sorting of frequent one itemset will be done using descending order and again it will be stored in a file F-List. By using this F-List signature matrix will be generated. Using k pivots similar transactions are grouped together and form a Group List with the corresponding unique group id (Gid). From the signature matrix LSH-Based partitioning is carried with corresponding Gid.

Reducer

Each reducer performs the local fp growth. Second mapreduce output will be used to generate frequent all patterns. fp growth while processing produce more dependent transactions are partition and it discovers the frequent transaction as final result.

C. Third MapReduce job

The last MapReduce job aggregates the result obtained from the second MapReduce job. Using the frequent pattern and generates candidate itemset for each itemsets and creates the final output.

Algorithm 1 LSH-fp-growth

Input: FList, k pivots, DB_i;
Output: Transactions for each Group id

1: MAP function contain(key as offset and values as $Database_i$)

2:   load FrequentList(FList), m pivots;

3:   Grouplist will be generated from the FList and m pivots

4:   for each Transaction in Database perform

5:   split the items in transaction and store it as items[]

6:   for every item in items[] perform

7:   if items is present in FList then store the item in b[]

8:   if end

9:   for end

10: put the Generated-signature matrix(b[]) into arraylist of signaturematrix;

11: For end

12: For every ((column)$_i$ in signaturematrix) perform

13: divide column of the signaturematrix into B bands with r rows;

14: Store the column in Hashbucket using hashmap

15: For end

16: if one band of column in signaturematrix and pivot p_k is mapped in to same bucket do

17: Assign j to the G_{id};

18: Output contains G_{id} and new transaction tree of b_i

19: If End

20: For each Grouplist (transaction≠i) do

21: if column in sigmatrix contains item in Grouplist then

22: Assign G_{id} to t;

23: Output contains G_{id} and new transaction tree of b_i

24: If end

25: For end

26: Function end

Input: transactions corresponding to each Gid;
Output: frequent k-itemsets

28: Reduce contains key as Groupid and values as database

29: Load Grouplists;

30: perform Grouplist according to Gid

31: perform local fp growth

32: for all (Transaction in Database G_{id}) do

33: build FP tree for the transaction

34: for end

35: for all element in b do

36: Define maxheap with the size K

37: perform TopKFpgrowth in maxheap

38: for all transaction in TopKFpgrowth

39: Output contains final transaction and support

40: for end

41: for end

42: function end

In LSH Algorithm each map phase takes transaction as pair <Long writable offset, Text record> then FList and k pivots are loaded initially. Each item in the transactions is split and compares with the FList items and stored in an array list b. From the b characteristic matrix and signature matrix will be generated. The signature matrix will be divide in to B bands each of which contains r rows (B x r= l). Then the column in the signature matrix are inserted in to the number of hash bucket if the sets are similar then it will be hashed to the same bucket. Assume that the similarity between two columns of signature matrix is d then they both transactions are considered similar. If a column in the signature matrix shares the similar bucket with a band B in pivot p then it will be denoted as a pair< p_k, T_i>. Finally the map output will be a pair of < p_k, T_i> and in reducer fp growth will be performed to generate frequent patterns.

ALGORITHM 2 Creation of SIGNATURE-MATRIX

Input: item transaction matrix of b[];
Output: generated signature matrix b[]

1:   Function to generate signature matrix of b[]

2:   For N number of hashfunction

3:   Assign every value in MinHash to max integer;

4:   For end

5:   For each(j=0;j<numberOfhashfunction;j++) perform

6:   For every element in the b[ ] perform

7:   Convert element to integer and store in value

8: Convert byte to hash of value and perform left shift 24 <- bytetohash[1];

9: Convert byte to hash of value and perform left shift 16 <- bytetohash[2];

10: Convert byte to hash of value and perform left shift 8 <- bytetohash[3];

11: Convert byte to hash of value <-bytetohash[4];

12: hashindex←hashfunction[j]*hash(bytesToHash);

13: if(minhashvalue[j]) > hashindex then

14: minhashvalues[j]=hashindex;

15: if end

16: For end

17: For end

18: Function end

Algorithm 2 describes about the generation of signature matrix. It is really difficult to create signature matrix from the large characteristic matrix. In order to avoid the complexity Minwise Independent permutation technique [10] is used to speed up the process. So the high dimensional matrix will be reduced to low dimensional matrix and time complexity is greatly reduced.

V.   DATASET DESCRIPTION

We generated a synthentic dataset using IBM quest market-basket data generator [11], initially the average transaction is set as 8 after that vary the average transaction and number of items to meet the various need in the experiment. The characteristic of our dataset are described in Table1.

Table 1:DataSet

Average length	Item in transaction	Average size of the transaction
8	1000	6.5B
10	4000	17.5B

40	10000	31.5B
60	10000	43.6B
85	10000	63.7B

VI. Experimental Evaluation

We evaluate the performance of partitioning technique in Hadoop equipped with 1 name node and 8 data nodes. Each node has an Intel Pentium-4.3.0 GHz processor 256 MB main memory and runs on the Linux 2.3 OS, on which java JDK 1.7.0 and Hadoop 2.1 are installed. We use the default replication factor 3 as a hadoop configuration parameter.

VII.   PERFORMANCE EVALUATION

7.1 Impact of pivot Number

We vary the number of pivot point from 20 to 100 and obtain the running time. Our partition technique achieve overall performance when compared to pfp technique. if we increase pivot number then Running time get increased. When the pivot number is 60 the running time is minimized.
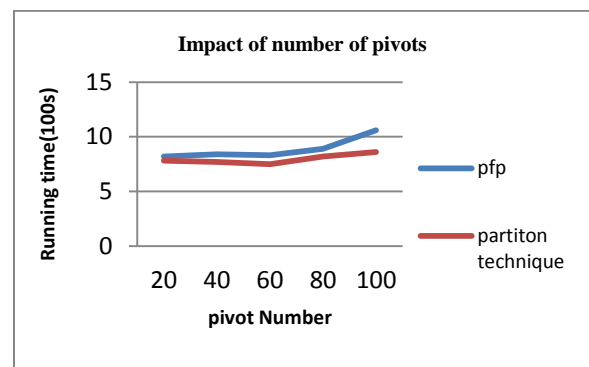
Figure 3. Impact of number of pivots

7.2 Impact of minimum support count

Figure 4. Impact of minimum support count

VIII. REFERENCE

[1] Yaling Xun, Jifu Zhang,Xiao Qin," FiDoop-DP:Data Partitioning in frequent itemset mining on Hadoop Clusters IEEE Transcations on Parallel and distributed system, vol28, jan.2017.

[2] M. J. Zaki, "Parallel and distributed association mining: A survey," IEEE Concurrency, vol. 7, no. 4, pp. 14–25, Oct. 1999.

[3] Pramudiono and M. Kitsuregawa, "Fp-tax: Tree structure based generalized association rule mining," in Proc. 9th ACM SIGMOD Workshop Res. Issues Data Mining Knowl. Discovery, 2004, pp. 60–63.

[4] M.-Y. Lin, P.-Y. Lee, and S.-C. Hsueh, "Apriori-based frequent itemset mining algorithms on mapreduce," in Proc. 6th Int. Conf. Ubiquitous Inform. Manag. Commun., 2012, pp. 76:1–76:8.

[5] X. Lin, "Mr-apriori: Association rules algorithm based on mapreduce," in Proc. IEEE 5th Int. Conf. Softw. Eng. Serv. Sci., 2014, pp. 141–144.

[6] S. Hong, Z. Huaxuan, C. Shiping, and H. Chunyan, "The study of improved FP-growth algorithm in mapreduce," in Proc. 1st Int.Workshop Cloud Comput. Inform. Security, 2013, pp. 250–253.

[7] M. Liroz-Gistau, R. Akbarinia, D. Agrawal, E. Pacitti, and P. Valduriez, "Data partitioning for minimizing transferred data in mapreduce," in

Proc. 6th Int. Conf. Data Manag. Cloud, Grid P2P Syst., 2013, pp. 1–12.

[8] L. Zhou, Z. Zhong, J. Chang, J. Li, J. Huang, and S. Feng, "Balanced parallel FP-growth with mapreduce," in Proc. IEEE Youth Conf. Inform. Comput. Telecommun., 2010, pp. 243–246.

[9] W. Lu, Y. Shen, S. Chen, and B. C. Ooi, "Efficient processing of k nearest neighbor joins using mapreduce," Proc. VLDB Endowment, vol. 5, no. 10, pp. 1016–1027, 2012.

[10] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining Massive Datasets. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[11] Z.Broder,M. Charikar,A. M. Frieze, and M. Mitzenmacher, "Min-wise independent permutations," J. Comput. Syst. Sci., vol. 60, no. 3, pp. 630– 659, 2000.

[12] L. Christopher. (2001). Artool Project [J].[Online].Available http://www.cs.umb.edu/laur/ARtool/ accessed Oct. 19, 2012

1st International Conference on Applied Soft Computing Techniques
22 & 23.04.2017
In association with
International Journal of Scientific Research in Science and Technology

A Comparative Study of TIWT and Shearlet Transform with Hard Thresholding for Normal Images

Syed Ali Fathima KMN*[1], Shajun Nisha S[2]
*[1]M.Phil(PG Scholar)PG & Research Dept of Computer Science, Sadakathullah Appa College,Tirunelveli,India
E-mail-Id:syedalifathima639@gmail.com[1]
[2]Prof.& Head,PG Dept of Computer Science,Sadakathullah Appa College, Tirunelveli,India
E-mail Id:shajunnisha_s@yahoo.com

ABSTRACT

Digital Images are generally corrupted by noise, Noise is nothing but addition of unwanted information for the Original Image. Image clatter is arbitrary discrepancy of luster or blush information in images, Removal of the noise is necessary to reduce the minimal damage of the image, improve image details. This paper describes a comparison of the discerning power of the different multimotion based thresholding techniques i.e., TIWT, Shearlet for image denoising. Shearlets are a multischematic structure which allows to efficiently encode anistropic features in multi types of various classes. Shearlet is a novel denoising method which can preserve edges efficiently. Translation invariant method improved the wavelet thresholding methods by averaging the estimation of all rendition of the degraded image. Inference of images which are denoised and its contrary problems, thus the experiments and conjectural analysis happen together. Comparatively the better evaluation of the result to produce shearlet transform.
Keywords: Denoising, TIWT Transform, Shearlet Transform, Hard Thresholding

I. INTRODUCTION

Image restoration is the removal or diminution of ruin that are incur ed while the image is being obtained. Degradation comes from smearing as well as noise due to electronic and photometric basis [1]. Image denoising is an important image processing task, both as a process itself, and as a constituent in other processes. Very many ways to denoised an image or a set of information exists. The main properties of a good image denoising model is that it will remove noise while defend edges. The goal of image denoising is to recuperate the true original image from such a indistinct piercing copy. The refurbish image should contain less noise than the interpretation while still observance spiky conversion (*i.e.* edges)[13]. Image denoising is a fundamental step in the image processing. Noise can distort by different inherent and exherent conditions. Depending on the devices noise may be additive and multiplicative noises. Additive noises are Gaussian and Salt and Pepper noise and Multiplicative introduced with the expressed intent to provide a highly proficient depiction of images with

edges. The elements of the shearlet representation form a collection of well-restricted waveforms, series at diverse positions, balances and directions and with highly anisotropic shapes[2].One of the innovative Wavelet Transform method is Translation-Invariant were introduced, Translation invariant method enhanced the wavelet thresholding methods by averaging the estimation of all transformations of the degraded image. A new de-noising method enhanced hard thresholding with the translation-invariant(Tl) wavelet transform is proposed in this paper.A translation Invariant wavelet transform is employed by exclude the sub-sampling at the each stage of the transform. Invariance means that you can identify an object as an object, even when its facade varies in some way[10]. This is generally a good thing, because it allows to extract an object's identity or category from the specifics of the visual input, like relative positions of the viewer/camera and the object). Wavelet transform (WT) is the most well-known two breadths and multi-ruling convert that decompose an image in horizontal, vertical and diagonal instructions[4].Researchers attempt to find new two

dimensions and multi-resolution transforms as the novel WT with more directionality in disparity with TIWT[19].Shearlet representation predominantly well acclimatized at representing the edges and the other anisotropic objects which are the prevailing features in typical images[6]. Recently, Labate *etc.*, described a new class of multi measuremental representation systems, called Shearlet, One advantage of this approach is that these systems can be erect using a generalized multi-pledge analysis and employed efficiently.   Hard threshold denoising method of Shearlet transform can get good concert, for its multi-degrees and multi-direction characteristic, image bare representation[9].Hence edges in an image get distorted. Shearlet transformation is a sparse, multi degree and multidimensional unconventional to wavelet transform. Shearlet Transform mingles multi extent and multi-trend demonstration and is very efficient to incarcerate inherent geometry of the multidimensional image and is optimally bare in representing image restraining edges[4]. Shearlets were introduced with the articulated purpose to provide a vastly competent representation of images with boundaries. In fact, the rudiments of the shearlet representation form a collection of well-contained waveforms, collection at different positions, scales and directions, and with highly anisotropic contours[5]. This makes the shearlet correspond to a particularly well acclimatized at representing the edges and the other anisotropic points which are the prevailing elements in archetypal images[2].The results appraise the recital of proposed filter and measure peak signal noise ratio. In this Paper intend, Shearlet is a narrative denoising method which can preserve edges competently better eradicates the noise from edges and without deforming the features[13].

Denoising Procedure:

The procedure to restored an image is given as follows:
De-noised image = W−1 [T{W (Original Image + Noise)}
Step 1: Apply ahead TIWT and Shearlet transform to a din image to get rotted an image.
Step 2: Apply hard thresholding to rotted image to eliminate noise.
Step 3: Apply inverse TIWT and Shearlet transform to thresholded image to get a denoised rotted an image[6].
Shearlet denoising With hard thresholding absorbs three basic steps- first step involves computation the Shearlet transform of noisy image, second step is used to apply thresholding on noisy Shearlet coefficient according to some rule and finally computing inverse Shearlet transform of amended Shearlet coefficients[15].

A.RELATED WORKS:

The deformations of images by noise are frequent during its acquisition, processing, compression, transmission, and re-production[6].The interfering throughout the conduction degrade the information. Noise may be generated by the transmission error and compression[13].In many applications, image denoising is used to produce good approximates of the original image from noisy observations[12].Translation Invariant denoising restrains noise by averaging over thresholded signals of all circular shifts. The hard thresholding will destroy all the coefficients whose enormity are less than the threshold to zero while keeping the continuing coefficients un amend. All the coefficients whose levels are greater than the threshold will be shrinked by the amount of the threshold[14].The denoising performance of wavelet transform methods is concerned by the following:

- Wavelet sources
- Number of disintegrations
- Transform type (orthogonal, outmoded, translation invariant, etc)

De-noising of normal images damaged by noise using wavelet techniques is very consequenced because of its ability to confine the force of a signal in few energy renovate values[11].The shearlet representation has appear in recent years as one of the most valuable constructions for the scrutiny and progression of multidimensional data[1]. Shearlet Transform which is based on the directional multiscale framework of the shearlet representation. In the field of Normal imaging, denoising requests to determine pertinent information in the several field as the shape, contour, etc. The Shearlet transform is useful for the noisy image to construct decomposed image coefficients[2]The scope of the paper is to focus on noise amputation techniques for normal images. Hard thresholding techniques are used for intention of image denoising. Keep and kill

rule which is not only automatically pleading but also initiates relics in the improved images is the basis of hard thresholding[16].In particular, the inaccuracy rates of data inference from noise are highly reliant on the sparsity possessions of the depiction, so that many successful applications of shearlets center around restoration chores such as denoising and contrary problems. Simple threshold denoising method of Shearlet transform can get good performance, for its multi-scale and multi-direction characteristic, image meager representation. However, there is a lot to be recovered.[4].The shearlet representation is a multi scale pyramid of well-confined waveforms delineated at various positions and directions, which was initiated to overcome the constraints of habitual multi scale structures in dealing with multidimensional data[4].Shearlet transformation is a bare, multiscale and multidimensional unconventional to wavelet transform. Shearlet transform is optimal in representing image containing edges[5].The results evaluate the performance of proposed filter and measure peak signal noise ratio. In this Paper Proposed, Shearlet is a new denoising technique which can conserve boundaries well-organized better eradicates the noise from edges and without warping the features[13].

B. MOTIVATION AND JUSITIFICATION

Image Denoising is the efficient one to denoise the images as well as to prevents the edges also. Removing or reducing noises from image is very important task in image processing. Image Denoising is utilized to develop and conserve the fine details that may be secreted in the data. In Image processing, noise is not easily eradicated as well as defend edges is also intricate. Shearlet is the greatest method for preserving the edges. Translation Invariant denoising smothers noise by middling more thresholded signals of all circular shifts. The hard thresholding will kill all the coefficients whose enormity are less than the threshold to zero while keeping the continuing coefficients unchanged. Shearlet Transform coalesces multiscale and multi-directional significations and is very competent to incarcerate inherent geometry of the multidimensional image and is optimally bare in signifying image restraining edges, The intend method using shearlet transform can be concerned to different types of normal images such as Lena, Cameraman, the shearlet transform can be implemented. Shearlet is best because it has

retained the precise information. During these advantages I motivate and justified to do work in Comparative transform such as(TIWT and Shearlet).

C. ORGANISATION OF THE PAPER:

This paper is preparation as follows. Methodology which contains the summarizes of the work of the TIWT and Shearlet Thresholding techniques are current in Section II. Experimental results are shown in Section III. Performance Analysis is also discussed in Section IV. finally Conclusion is presented in Section V.

II. METHODOLOGY

A.OUTLINE OF THE WORK :

Denoising utilizes TIWT, Shearlet Transform. This system is eloquented in Fig 1.Gaussian and Salt & Poisson nosie is added with the two enter images such as Lena and Cameraman. TIWT, Shearlet Transform is used to crumbled the noisy image and then apply the Hard threshold function to the noisy image to get the denoised output image.
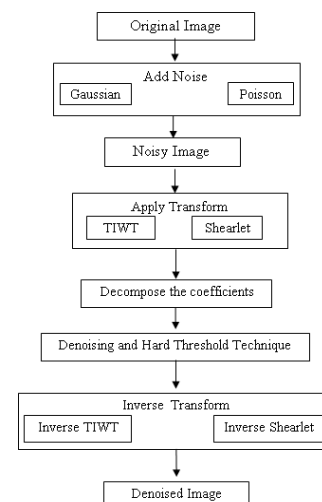
Figure 1. Block Diagram for TIWT and Shearlet with Hard Threshold

B.TIWT, SHEARLET TRANSFORM

TI wavelet denoising suppresses noise by averaging over thresholded signals of all spherical shifts. TI wavelet Denoising better than the traditional wavelet transform may generate relics on discontinuities of the signal. the de-noised signal can be renovated using inverse wavelet[20].Pseudo-Gibbs phenomena are connected with

succession position of original signal singularities. So Coifman and Donoho[21] locate TI de-noising to eradicate this defect.

Given a signal $x(n)(n=0,1,…,N-1)$,defined *Sh* as arithmetic operator of rendition[15], the value of reallocating is h: $Sh(x(n)) = x((n+h) \bmod N)$
For all x, y a $\in$ X and every scalar α

$$d(x+a, y+a) = ||x+a-(y+a)|| = ||x-y|| = d(x,y)$$
$$d(\alpha x, \alpha y) = ||\alpha x - \alpha y|| = |\alpha| ||x-y|| = |\alpha| d(x,y)$$
$$||x||_1 = |\xi_1| + |\xi_2| \quad ||x||_p = (\xi_1^p + \xi_2^p)^{1/p}$$
$$||x||_2 = (\xi_1^2 + \xi_2^2)^{1/2} \quad ||x||_\infty = \max\{|\xi_1|, |\xi_2|\}$$

Shearlet have well restricted waveforms and soaring bearing sensitivity contrast to other state-of-art procedures. Shearlets are correlated with multi scale and multidirectional decomposition, which enable them to capture intrinsic geometric features of image.

Shearlet has high directional sensitivity and are optimally meager in corresponding image containing edges. Shearlets are constructed by parabolic Scaling Fig.2, shearing and translation applied.
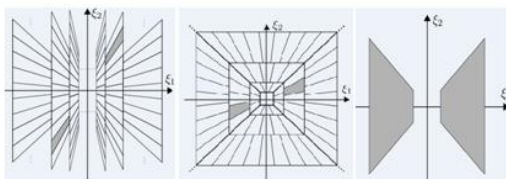
Figure 2. Shearlet Transform

As noise has a fine crumbed structure in the image therefore most of the noise Shearlet Transform is represented here, $L^2(R^2)$.
Compared the two transform involves three basic steps are as follows:
i) first step rivets calculation the TIWT and Shearlet transform of noisy image,
ii) second step is make use to apply thresholding on noisy TIWT and Shearlet coefficients concurrence to some rule and
iii) finally calculate inverse TIWT and Shearlet transform of modified these coefficients.

C. HARD THRESHOLDING FUCNTION(TIWT AND SHEARLET)

Thresholding is a procedure used for signal and image denoising. The hard-thresholding function desires all Shearlet coefficients that are greater than the provide threshold λ and sets the others to zero. The threshold λ is chosen according to the signal energy[6].These methods are to make a noises free in an image.

$$D(x,y) = \begin{cases} S(x,y) & if\ S(x,y) > T \\ else & 0 \end{cases}$$

Where T is Hard threshold. Let S(x, y) represent the original shearlet coefficient in the point (x, y) in each sub-band K $\in$ { K1 K2 …. Kj} at scale j. The intend of this paper is to obtain denoised coefficient D(x,y) at the position S(x,y) by adjust the pixel values.

Step 1: wavelet decomposition of the image: establish the wavelet function and decomposition levels *N* , and decompose the image with *N* layer wavelet.
Step 2: Threshold selection: choose the threshold for each TI wavelet coefficients of each layer, and critic the threshold of detail coefficients.
Step 3: TIWT coefficient with threshold progression will be used to reform the image by inverse wavelet transform[10].
R. Coifman and D. Donoho[21] improved the wavelet thresholding methods by averaging the calculate approximately on of all transformations of the degraded image.

D. DECOMPOSE THE COEFFICIENTS:

Normally, those wavelet coefficients with diminutive degrees than the preset threshold are caused by the noise and reinstated by zero, and the others with better magnitudes than the preset threshold are caused by original signal, Then the denoised signal could be re-enacted from the resulting wavelet coefficients[21].A new translation-invariant (TI) wavelet denoising method with developed hard thresholding is presented to reduce the noise from the denoised image.

Figure 3. Lena image two level Decomposition

Figure 4. Cameraman image two level Decomposition

The shearlet decomposition procedure is instigated by detaching the image into its high pass and low pass constituents, Shearlet decomposition results in large number of shearlet coefficients and we need to detach noisy coefficients from original ones. Thresholding is very important because thresholding at outsized values result in beating of information whereas at stumpy values result in backgr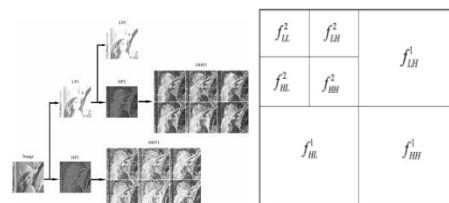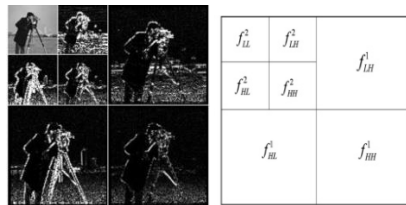ound encumber. A two level shearlet decomposition wherein each level consisting of 3, 3, 4 and 4 numbers of trimming directions respectively. Thus, the number of directional sub bands within each level was obtained as 8,8,16 and 16 respectively as the number of directional sub-bands within each level $N_s = 2^s$ where N_s is the number of shearing directions[11].

The process, consists of following main stages:

1) comprehend the noisy image as input
2) Perform TIWT and shearlet of noisy image and obtain TIWT and Shearlet coefficients
3) calculate approximately noise variance from noisy image
4) determine threshold value using various threshold selection statutes.
5) Apply hard thresholding function to noisy coefficients
6) Perform the inverse TIWT and shearlet to re-enact the denoised image[16].

E.NOISE CATEGORIES

Digital images are often damaged by many types of noise comprising salt and pepper noise, Gaussian noise, Poisson noise, Depending on the type of interruption; the noise can involve the image to dissimilar area. normally our focal point is to confiscate certain kind of noise. Noise is the unwanted effects generated in the image, Image noise can be categorized as Impulse noise (Salt-and-pepper noise), Amplifier noise (Gaussian noise), Shot noise, Quantization noise (uniform noise), Film grain, on-isotropic noise, Multiplicative noise (Speckle noise) and Periodic noise[14].

A. Gaussian Noise (Amplifier Noise)

The term normal noise model is the synonym of Gaussian noise. This noise model is additive in nature [4] and follow Gaussian distribution. Meaning that each pixel in the noisy image is the sum of the true pixel value and a random, Gaussian distributed noise value. The noise is independent of intensity of pixel value at each point. *P(z)* is the Gaussian distribution noise in image; μ and σ is the mean and standard deviation respectively.

$$p(z) = \frac{\frac{1}{\sqrt{2\pi\sigma}} e^{-(z-\mu)^2}}{2\sigma^2}$$

B. Poisson Noise (Photon Noise)

Poisson or shot photon noise is the noise that can reason, when number of photons intellect by the sensor is not ample to give evident numerical information[4]. This noise has root mean square value comparative to square root intensity of the image. Different pixels are beared by sovereign noise values. At convenient grounds the photon noise and other sensor pedestal noise distort the signal at different proportions.

$$p(x) = e^{-\lambda} \lambda^x \quad \text{for } \lambda > 0 \text{ and } x = 0,1,2....$$

III.EXPERIMENTAL RESULTS
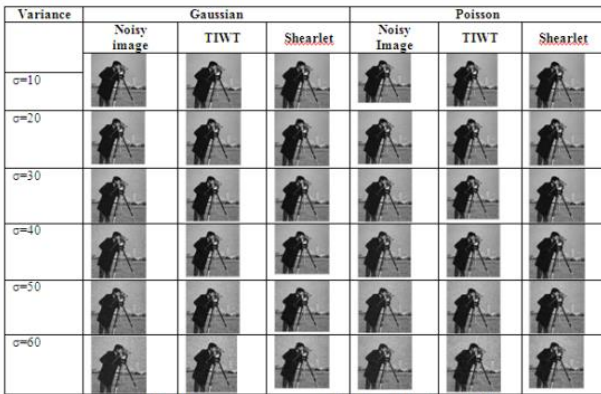

Figure 4. Original Image for Lena and Cameraman

Figure 5. Lena image using TIWT and shearlet with hard threshold

Figure 6. Cameraman image using TIWT and shearlet with hard threshold

IV. PERFORMANCE ANALYSIS

A. PERFORMANCE PARAMETERS:

i) PSNR:

PSNR is used to appraise the restoration results, which determines how secure the restored image is to the original image. It is the ratio between maximum achievable power of a signal and the power of fraudulent noise that distress the superiority and dependability of its representation. PSNR is calculated as,

$$PSNR = 20 * log \log_{10} \frac{maxi}{\sqrt{mse}}$$

ii) MSE:

The slighter the MSE the nearer the estimator is to the tangible data. A miniature mean squared error means that the arbitrariness reflects the data more precisely than a superior mean squared error. The goal is to estimation the signal x_{ij} from noisy observations yij such that Mean Squared error (MSE) is minimum. I.e.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i-j) - K(i-j)]^2$$

iii) RMS:

The RMS overall time of a sporadic function is equal to the RMS of one period of the function. The RMS importance of a incessant function signal can be inexact by intriguing the RMS of a cycle of equally spaced samples. Additionally, the RMS value of various waveforms can also be established,

$$x_{rms} = \sqrt{\frac{1}{n}(x_{1^2} + x_{2^2} + x_{3^2} + \cdots + x_{n^2})}$$

B .PERFORMANCE EVALUATION

The Recital of TIWT and Shearlet was calculated by using MSE,RMS,PSNR. The experimentation of TIWT and Shearlet with hard threshold Gaussian noise have been calculated and their denoised image results are shown in Table I, and Table II. Measured all the metrics, it is scrutinized that the TIWT and Shearlet presents well for Normal images such as Lena and Cameraman. In Table I, II analysis it identify the best threshold for a particular noises. From Table I, its observed that Hard Threshold is well Suitable for both Gaussian and Poisson noise. In Table II Poisson noise is better amputation for Hard threshold and Poisson noise is better elimination for Transforms.

TABLE I TIWT AND SHEARLET FOR DENOISED NORMAL IMAGES GAUSSIAN NOISE

Image	Variance	TI by hard threshold			Shearlet by hard threshold			Time by TI	Time by Shearlet
		MSE	RMS	PSNR	MSE	RMS	PSNR		
Lena	σ=10	07.90	02.81	30.17	07.34	02.71	30.82	09.85	5.81
	σ=20	07.96	02.82	30.11	07.35	02.71	30.80	09.08	5.72
	σ=30	08.64	02.93	29.40	07.78	02.79	30.31	09.10	5.73
	σ=40	13.60	03.68	25.45	11.37	03.37	27.01	09.15	5.78
	σ=50	24.08	04.90	20.49	20.09	04.48	22.07	09.11	5.74
	σ=60	36.41	06.03	16.90	31.68	05.62	18.11	09.14	5.83
Cameraman	σ=10	07.35	02.71	30.80	06.91	02.62	31.34	11.87	5.77
	σ=20	07.47	02.73	30.66	06.99	02.64	31.24	10.37	5.83
	σ=30	08.19	02.86	29.86	07.46	02.73	30.67	09.92	5.91
	σ=40	13.59	03.68	25.47	11.25	03.35	27.10	12.21	5.79
	σ=50	24.12	04.91	20.48	20.16	04.49	22.04	09.87	5.91
	σ=60	36.43	06.03	16.90	31.81	05.64	18.08	10.66	5.77

TABLE II TIWT AND SHEARLET FOR DENOISED NORMAL IMAGES POISSON NOISE

Image	Variance	TI by hard threshold			Shearlet by hard threshold			Time by TI	Time by Shearlet
		MSE	RMS	PSNR	MSE	RMS	PSNR		
Lena	σ=10	07.56	02.63	31.87	06.94	02.37	32.12	09.12	05.61
	σ=20	07.85	02.55	31.09	07.12	02.39	31.82	08.85	05.43
	σ=30	08.52	02.79	30.41	07.18	02.54	21.23	09.00	05.52
	σ=40	13.23	03.52	28.65	11.13	03.12	29.95	09.04	05.58
	σ=50	23.86	03.94	25.79	19.92	03.78	27.53	09.09	05.46
	σ=60	35.31	05.93	22.90	30.86	04.54	24.82	09.07	05.37
Cameraman	σ=10	07.12	02.59	31.93	06.72	02.21	32.37	10.82	05.23
	σ=20	07.27	02.66	31.65	06.32	02.57	31.93	10.15	05.73
	σ=30	08.03	02.72	30.68	07.15	02.78	31.72	09.80	05.76
	σ=40	12.78	03.38	28.74	10.75	03.27	29.19	12.00	05.62
	σ=50	24.01	04.73	23.68	19.21	04.23	23.58	09.33	05.31
	σ=60	35.96	05.93	19.79	30.37	04.92	20.65	10.43	05.27

IV. CONCLUSION

This paper present for Normal image denoising pedestal on TIWT and Shearlet. shearlet is one of the preeminent method for decomposition compared with the TIWT. The images are damaged with Gaussian and Poisson noises with Hard Thresholding technique. After denoising, the image transparency is too improved. Quantitative performance determine such as MSE,RMS,PSNR are used to estimated the denoised image achieve. From the result it is observed that

Gaussian noise and Poisson are well appropriate for Hard Thresholding Functions. Shearlet could propose superior results than hard thresholding to get better the finest excellence of the image.

III. REFERENCES

[1] Miss Monika shukla1, Dr.Soni changlani2," A Comparative Study of Wavelet and Curvelet Transform for Image Denoising", changlani2IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735. Volume 7, Issue 4 (Sep. - Oct. 2013), PP 63-68

[2] Mr. Rohit Verma1 Dr. Jahid Ali2, "A Comparative Study of Various Types of Image Noise and Efficient Noise Removal Techniques",International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014)

[3] Glenn R. Easley1 and Demetrio Labate2," Image Processing using Shearlets"

[4] Zhiyong Fan1,2, Quansen Sun1, Feng Ruan2, Yiguang Gong2 and Zexuan Ji1 ," An Improved Image Denoising Algorithm based on Shearlet ",International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 4, August, 2013

[5] Pooran Singh Negi1 and Demetrio Labate2," 3D Discrete Shearlet Transform and Video Processing"

[6] G. R. Easley, D. Labate, and W.Q. Lim,"Sparse directional image representations using the discrete shearlet transform,"Appl. Comput. Harmon. Analysis, vol.25, Jan. 2008, pp.25-46.

[7] Anju T S1, Mr. Nelwin Raj N R2 " Shearlet Transform Based Satellite Image Denoising Using Optimized Otsu Threshold", International Journal of Latest Trends in Engineering and Technology (IJLTET)

[8] "Denosing ECG using Translation Invariant Multiwavelet", World Academy of Science, Engineering and Technology , Jeong Yup Han, Su Kyung Lee, and Hong Bae ParkInternational Journal of Electrical, Computer, Energetic,

Electronic and Communication Engineering Vol:2, No:5, 2008.

[9] Rajesh Kumar Rai1, Trimbak R. Sontakke2,"Implementation of Image Denoising using Thresholding Techniques", ISSN 2249-6343 International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1 , Issue 2. Trends in Engineering and Technology (IJARTET) Vol.3, Issue4,April'16

[10] Mrs.P.Sivamani1,Mrs.V.VidhyaGowri2,Ms.S.V. Priyavarshini3, Ms.N.Revathi4, "Image Denoising Using Wavelet Thresholding", International Journal of Advanced Research journal of computer Technology and Electronics Engg(IJCTEE)

[11] Shubh Karman Kaur1, Rupinder Kaur2,"An Efficient Threshold Based Mixed Noise Removal Technique",International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue 7 308 – 311.

[12] Arun Dixit, Poonam Sharma,"A Comparative Study of Wavelet Thresholding for Image Denoising",I.J. Image, Graphics and Signal Processing, 2014, 12, 39-46 Published Online November 2014 in MECS DOI: 10.5815/ijigsp.2014.12.06

[13] R. Bouchouareb and D. Benatia,"Comparative Study between Wavelet Thresholding Techniques (Hard, Soft and Invariant-translation) in Ultrasound Images", international Journal of Bio-Science and Bio-Technology Vol.6, No. 6 (2014), pp.29-38.

[14] "An Effective Signal De-noising Algorithm Combining Optimal Wavelet Packet Basis and Translation-Invariant", Di Wu Zhejiang University,Conference Paper,June 2008.

[15] Syed Ali Fathima KMN1, Shajun Nisha S2," Shearlet Transform Based Normal Image Denoising Using Hard Threshold", International Journal of Innovative Research in Computer and Communication Engineering,Vol. 4, Issue 11, November 2016.

[16] "Translation Invariant Wavelet Transform Based Image Denoising on Normal images using Hard Threshold",Syed Ali Fathima KMN1and Shajun

Nisha2,"Attend in National Conference in "Recend Trends in Data Minining at Xavier's College,Tirunelveli.

[17] Jannath Firthouse P1, Latha Rani G.L2, Shajun Nisha S3,"An Effective Shrinkage Threshold for Contourlet based on Image Denoising in Natural Images,International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 6, June 2016.

[18] Anutam1 and Rajni2"Performance Analysis Of Image Denoising with Wavelet Thresholding Methods for Different levels of Decomposition", International Journal Of Multimedia & Its Applications (IJMA) vol.6, no.3, June 2014

[19] Abdullah Al Jumah,"Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Techniques", Journal of Signal and Information Processing, 2013, 4, 33-41.

[20] "The denoise based on translation invariance wavelet transform and its applications", Shuren Qin, Changqi Yang, Tang Baoping and Shanwen Tan, The Test Center, Institute of Mechanical Engineering, Chongqing University Chongqing 400044, P. R. China.

BIOGRAPHY

K.M.N.Syed Ali Fathima received the B.Sc degree in Computer Science from MS University in 2012 and M.Sc degree in Computer Science from MS University in 2015. She is currently pursuing the M.Phil degree in Computer Science under the guidance of S. Shajun Nisha. Her Research interest are mainly include domain of Image Denoising.

Shajun Nisha S, Professor and Head of the Department of Computer Science, Sadakathullah Appa College, Tirunelveli. She has completed M.Phil. (Computer Science) and M.Tech (Computer and Information Technology) in Manonmaniam Sundaranar University, Tirunelveli. She has involved in various academic activities. She has attended so many national and international seminars, conferences and presented numerous research papers. She is a member of ISTE and IEANG and her specialization is Image Mining.

Effective Retrieval and Analysis of Uropathogens through NoSQL Database

Dr. V.T. Meenatchi, Dr. M. Thangaraj

Department of CA & IT, Thiagarajar, College, Department of Computer Science, M.K University, Madurai, India

S. Padmavathy, N. K. AshaDevi, K. Vignesh

Department of Zoology & Microbiology, Thiagarajar College, Madurai, India

ABSTRACT

In the today's web era, big data is emerging. The storage and retrieval of big data is becoming an issue. The database administrators are moving into new storage technology, the NoSQL database. This paper analyzes the predominant organisms causing the Urinary Tract Infection (UTI) based on gender wise and age-wise and also the antibiogram pattern of 3G and 4G antibiotics were analysed. The work is clinically proven using new methodologies and the data is then mapped into MongoDB, a NoSQL database. Through this type of mapping and analysis, the data retrieval becomes ease and simpler to manage huge data. The generated analytical report aids the medical practitioners to provide the needful therapy for UTI affected patients.

Keywords : NoSQL; MongoDB; Urinary Tract Infection; Antibiotics; Therapy

I. INTRODUCTION

One of the most prevalent problems faced by healthcare services is the increasing prevalence of antimicrobial resistance. Urinary tract infections (UTI) are the most common bacterial infections affecting humans throughout their lifetime. UTI is a serious ailment in human due to increasing frequency, recurrence and difficulty in eradication; it poses stiff challenge to the medical professionals. It is much more common in women than in men, due to anatomical and physiological reasons by virtue of its position urinogenital tract is more vulnerable to bacterial infections caused by both internal and external flora [3].

Every year, the rate of people affected by urinary tract infections is increasing. when the infections are not treated, it leads to serious health problems. This work aims to identify the uropathogens causing the infection. From the urine specimens, it was noted that the following uropathogens like Escherichia coli, Klebsiella pneumoniae, Citrobacter sp., Enterobacter sp., and Staphylococcus aureus are responsible main cause for infection. The NOSQL databases were created as a mean to offer high performance and high availability. NOSQL

is an unstructured Query Language [11] and data stores are widely used to store and retrieve large amounts of data.

Our paper is organized as follows, Section II discusses the related literatures, Section III deals about the proposed system and Section IV presents the Results and Discussion and Section V ends with Conclusion.

II. LITERATURE SURVEY

Urinary tract infection is one of a serious health problem affecting millions of people yearly. Infections of urinary tract are in the second most common type of infection in the body. Urinary tract infection has an important association in human female, the highest incidence of urinary tract infection occur in child bearing age and this has been linked to sexual activity and aging. Asymptomatic urinary tract infection occurs in 2 to 10% in women during pregnancy.

Urinary tract infection can be seen in three different forms in infected pregnant women, asymptomatic bacteriuria, acutecystitis or acute pyelonephritis. The incidence of asymptomatic urinary tract infection has

been reported between 2-13% however physiological changes in pregnancy lead to severe course of problem and also it has been reported that if untreated asymptomatic bacteriuria increase the frequency of premature delivery and neonates with low birth weights[7]. Urinary tract infection (UTI) is one of the most common and life threatening infection present in community practice.

Manifestations can vary from asymptomatic bacteriuria to symptomatic cystitis, pyelonephritis and blood stream infection[8]. The most common symptoms are pain, fever, or nausea and vomiting in addition to the classic symptoms of a lower urinary tract infection. Rarely the urine may appear bloody[9]. The general symptoms include a frequent urge to urinate a painful burning feeling during urination. These urinary infection are known to cause to a lot of problems in the physiological action of the urinary system. There are a number of reasons why bacterial resistance should be a concern for physicians. First resistant bacteria particularly Staphylococci, Enterococci, Klebsiella pneumoniae and Pseudomonas sp. are becoming common place in healthcare institutions. Bacterial resistance often results in treatment failure, which antibacterial therapy, defined as the initial use of an antibacterial agent to which the causative pathogen was not susceptible, has been associated with increased mortality rates in patients with bloodstream infections due to resistant Pseudomonas aeruginosa, Staphylococcus aureus, Klebsiella pneumoniae, Escherichia coli, Enterobacter sp., coagulase negative Staphylococci and Enterococci. Prolonged therapy with antimicrobial agents such as vancomycin or linezolid, may also lead to the development of low – level resistance that compromises therapy, but that may not be detected by routine susceptibility testing methods used in hospital laboratories. Many uropathogen are producing the Beta lactamase enzymes, so drugs resistant to particular pathogen for example Hydrolysis of beta lactam ring by beta lactamase enzymes produce by uropathogen.

During cystitis, Uropathogenic Escherichia coli subvert innate defenses by invading superficial umbrella cells and rapidly increasing in numbers to form intracellular bacterial communities [2]. By working together, bacteria in biofilms build themselves into structures that are more firmly anchored in infected cells and are more resistant to immune-system assaults and antibiotic treatments.

This is often the cause of chronic urinary tract infections. Antibiotics are the main treatment for all UTIs. A variety of antibiotics are available, and choices depend on many factors, including whether the infection is complicated or uncomplicated or primary or recurrent.

Worldwide reports of antibiotic resistant E. coli isolates indicate the unwise and excessive consumption of antimicrobial drugs which in turn has brought about failure in treatment, and consequently concerns about the related issues in all nations including the developed and developing ones. As an example, TMP-SXT was conventionally used for uncomplicated cystitis in most regions, however, due to the resistance to it, fluoroquinolon and cephalosporin took its place and unfortunately after sometime resistance to these two drugs was also recognized and reported [4][5].

Reports of uropathogens resistant to previously effective antibiotics have emerged globally in recent years. The situation is especially dire in Africa where irrational antibiotic practices are common. Variations in antibiotic resistance patterns are known to occur across different geographical regions, even within the same country. Such variations must be well documented so as to inform local empirical treatment as well as foster rational antibiotic use. The treatment of bacterial infections is increasingly complicated by the ability of bacteria to develop resistance to antimicrobial agents [6].

The development of resistance to all available antibiotics in some organisms then precludes the effectiveness of any antibiotic region. Organisms that are resistant to all known effective antimicrobials pose a serious threat to hospitalized patients. Indiscriminate use of antibiotics leads to the development of resistance of initially sensitive strains of organisms and possibly destruction of the normal flora [1].

The traditional database system maintained are the relational database systems, in which all the database retrieval and transactions are done in (SQL). Since the database becomes huge in size, we have to retrieve databases through MongoDB [10]. The necessity of using MongoDB is to have document oriented storage. MongoDB is applied in Big Data, Content Management a Delivery, Mobile and Social Infrastructure, User Data Management, Data Hub. Database in MongoDB[12], is a physical container for collections. Each database gets its

own set of files on the file system. A single MongoDB server[13] typically has multiple databases.

III. PROPOSED SYSTEM

An architecture named as UTI_framework in Fig. 1 has been constructed with 6 main components namely,

- Sample collection
- clinical analysis with three subcomponents:
  - Morphological characterization
  - Culture characterization
  - Biochemical characterization
- Uropathogens Identification
- Mining through NoSQL
- Antibiogram Pattern
- Analytical Report
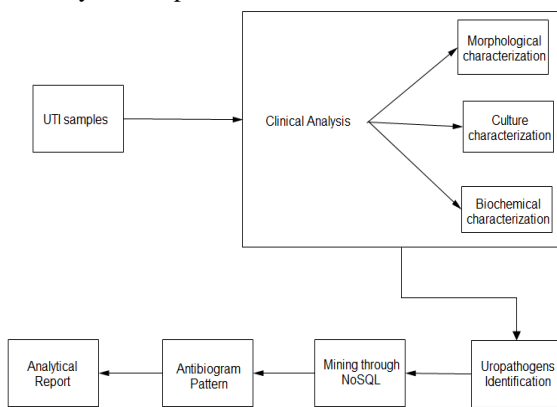
Figure 1. UTI Framework

All the components in the framework are interdependent. The collected urine samples were given for clinical analysis for uropathogen identification. The results were then mapped into MongoDB 3.2 document store environment for mining [14]. The Antibiogram patterns were assessed for specific organism. Finally the reports were generated to depict organism predominance, proper antibiotic for treatment, its sensitive level to organism, gender and age wise analysis.

IV. RESULTS AND DISCUSSION

Following are the results derived through clinical study. For analysis, 500 samples were collected and major affected were female than male as shown in Table 1.

TABLE I. Sex distribution of UTI Patients

Sex	Total number
Male	150
Female	350
Total	500

Various organisms were identified and number of uropathogens with their corresponding percentage is shown in Table 2.

TABLE II. Percentage of Isolated Uropathogens using Selective and Differential Medium

Name of the organisms	Number of Uropathogens	Percentage of Uropathogens
E.coli	88	31.10
K.pneumoniae	55	19.43
Staphylococcus aureus	42	14.84
Enterobacter sp.,	32	11.31
Pseudomonas sp.,	27	9.54
Streptococcussp.,	20	7.07
Citrobacter sp.,	13	4.59
Proteus sp.,	6	2.12
Infected	283	-
Non infected	217	-
Total	500	100

The age wise and gender wise analysis were made and shown in Table 3.
From the table, one could identify the maximum percentage affected in male falls in 50 years and above, whereas in case of female, the maximum percentage falls under 30-39 years age group.

TABLE III. Distribution of the UTI Cases due to *E.coli* with reference to Age and Gender Status

Age groups (in Years)	Male		Female	
	Total	Percentage	Total	Percentage
Less than 1	2	4.34	7	2.37
1-19	8	17.39	29	10.14
20-29	3	6.53	58	20.39
30-39	5	10.87	82	28.94
40-49	10	21.74	57	20.14
50 and above	18	39.13	51	18.02

The antibiogram pattern for the infected organisms with their sensitivity levels are depicted in Table 4.

TABLE IV. Antibiogram Pattern of Uropathogenic *E.coli* before Plasmid Curing

Antimicrobial Agents	Sensitivity Isolates	Intermediate Isolates	Resistant Isolates
Amikacin (30 µgms)	222	51	10
Nitrofurantoin (300 µgms)	185	55	43
Ampicillin/Sulbactam (10/10 µgms)	86	145	52
Kanamycin (30µgms)	40	165	78
Norfloxacin(10 µgms)	92	80	113
Meropenem (10 µgms)	260	15	8
Cefixime (5µgms)	37	79	167
Ciprofloxacin (5 µgms)	85	96	102
Ofloxacin (5µgms)	96	99	88
Co-Trimoxazole (23.75/1.25 µgms)	77	40	166
Pristinamycin (15µgms)	46	104	133
Vancomycin (30 µgms)	34	32	217
Gentamicin (10µgms)	53	195	35
Cefoperazone (75 µgms)	36	115	132
Chloramphenicol (30µgms)	177	51	55
Linezolid (30 µgms)	42	34	207
Azithromycin (15 µgms)	108	114	61
Erythromycin (15 µgms)	39	84	160
Piperacillin/Tazobactam(100/10µgms)	184	63	36
Sulphafurazole (300µgms)	65	57	161
Tobramycin (10 µgms)	40	215	28
Oxytetracycline (30µgms)	66	41	176
Clarithromycin (15 µgms)	24	86	173
Doxycycline Hydrochloride (30 µgms)	65	141	77
Moxifloxacin (5 µgms)	103	131	49

After clinical analysis, the results were mapped into MongoDB as shown in Fig. 2 , for effective retrieval.
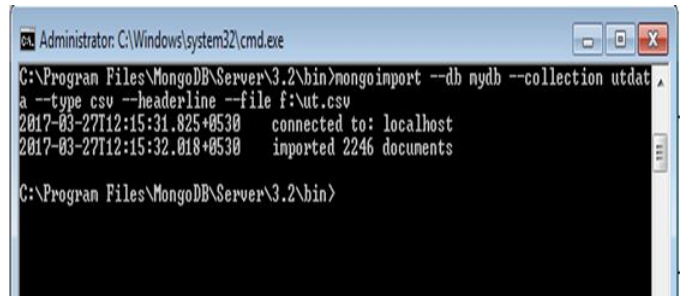

Figure 2. Mapping of UTI data into NoSQL

After mapping the data is depicted as shown in Fig. 3.From the UTI dataset, out of 200, 138 number of female were affected from Escherichia Coli as shown in Fig. 4.
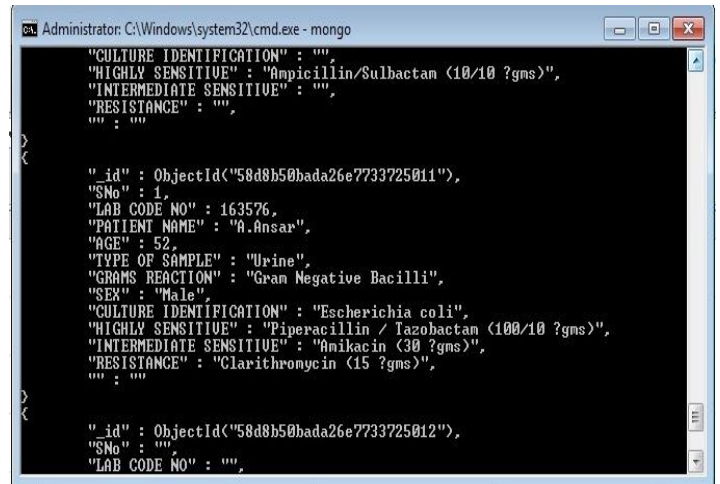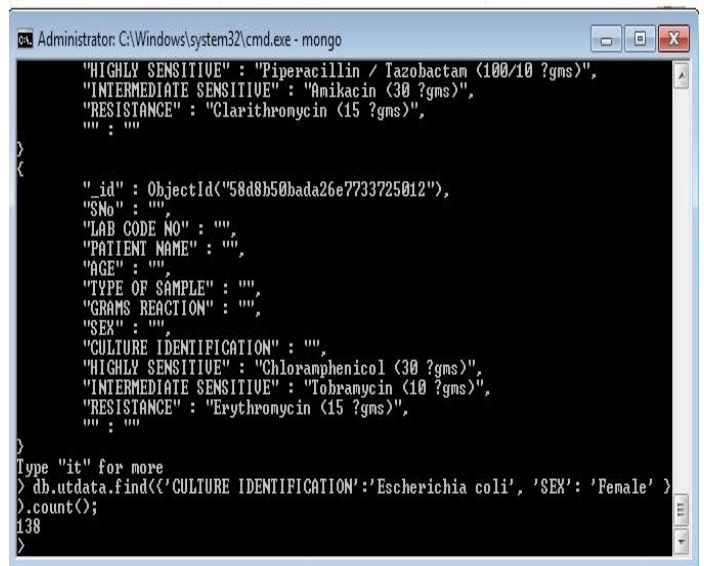

Figure 3. Display of documents

Figure 4. Culture Identification

Fig.5 shows the number of isolates affected through E. Coli from 100 samples.

Figure 5. E. Coli affected samples

Thus the desired reports can be generated through NoSQL database and any useful information can be tracked anytime when needed.

V. Conclusion

It is concluded from the present study that routine microbiological analysis for antibiotics sensitivity tests for urinary tract infected patients and other patients to be carried out before administration of drugs for treatment and management of urinary tract infections, since resistance to these drugs are developing in the community. This study appears to suggest a need for a continuous monitoring of bacterial antibiotics susceptibility before antibiotics prescription in order to ensure adequate treatment for urinary tract infection and reduction in the spread of bacteria resistant strain. Identification and proper treatment of UTI infections will lead to 10 fold decrease in the occurrence of acute pyelonephritis. After clinical study, data is computerized through NoSQL database. This makes the overall retrieval and mining easier. Moreover, self-medication should be avoided in order to prevent spread of drug resistant strains of bacteria.

VI. References

[1]   D.K.Chaturvedi, P.S.Satsangi , P.K. Kalra "Load frequency control:a generalized neural network approach" Electrical Power and Energy Systems 21(1999)405-415 Elsevier

[2]   Dola Gobinda Padhan,Somanath Majhi " A new control scheme for PID load frequency controller for single area and multi area systems" ISA Transactions 52(2013)242-251 Elsevier

[3]   O.I.Elgerd "Electric Energy System Theory;An introduction "Mc Graw Hill.1971

[4]   M.K.Sherbiny El "Efficient fuzzy logic load frequency controller"Energy Conversion and Management 43(2002)1853-1863 Elsevier

[5]   S.P.Ghosal "Optimizations of PID gains by particle swarm optimizations in fuzzy based automatic generation control"Electric Power Systems research 72(2004)203-212 Elsevier

[6]   Muwaffaq Irsheid Alomoush "Load frequency cntrol and automatic generation control using fractional order controllers" Electr Engg(2010)91:357-3688 Springer

[7]   Wen tan"tuning of PID load frequency controller for power systems" Energy Conversion and Management 50(2009(1465-1472 Elsevier

[8]   V.soni.G.Parmar,M.Kumar and S.Panda "hybrid grey wold optimization-pattern search optimized 2DOF-PID controllers for load frequency control in interconnected thermal power plants"ICTACT journal on soft computing April 2016,volume 06, issue 03.

[9]   RV Rao, VJ savsani,DP vakharia "Teaching learning based optimization: an optimization method for continuous non-linear large scale problems"Infinite science 2012;183(10:1-15

[10]   P. Kundur."Power system stability and control",Tata Mc Graw Hill,2009.

[11]   J.Talaq,A.I.Fadel and Basri,"Adaptive Fuzzy Gain Scheduling for Load Frequency control",IEEE Transactions on Power System,Vol.14.No.1,pp.145-150,1999

[12]   D.P.KothariandI. J. Nagrath." Power System Engineering", 2nd edition,Tata

Diminution of Testcases in Object Oriented Software using JUnit Tool

B.Geetha[*1], V.Bhaskar[2], Dr.D.Jeya Mala[3]
*[*1]Department of Master of Computer Applications, KLN College Of engineering, Madurai, Tamilnadu,India*
bgeet76@gmail.com[1]
[2]Department of Computer Science Engineering, KLN College Of Engineering, Madurai, Tamilnadu,India
bhaskarv@gmail.com[2]
[3]Department of Master Of computer Applications, Thiagarajar College Of Engineering, Madurai,Tamilnadu,India

ABSTRACT

To ensure the code is working the way the developer intends to test the software.Testing is very costliest phase in software development.The object oriented programming paradigm generating test data is very challenging one.To reach high code coverage search based software testing is used.This paper introduces a novel way to deal with decreases the quantity of experiments utilizing JUnit tool.It likewise tackles the requesting of articles utilizing reflexive strategy in question situated system.This approach comprises of three phases.The first stage distinguishes the kind of protest utilizing reflexive method.In this second stage we build up a need keeping in mind the end goal to keep the abnormality of testing order.In the third phase,we diminish the no of test cases in object oriented software using JUnit tool.

Keywords: Object Oriented, Junit Tool, Software Testing

I. INTRODUCTION

Test-information era is a basic and imperative phase of programming development.Manually producing test-information is exertion devouring task.Revealing concealed mistakes in code execution is exceptionally tedious around half time of the aggregate programming life-cycle[3].Object arranged programming testing is extremely testing task,because of its elements like conceptual class, Inheritance, polymorphism and perceivability.

Many methodologies are accessible to create programmed test cases which incorporate distinctive instruments and strategies which abridged as takes after, Code based test information era techniques.There are fundamentally 3 techniques used to produce test information, Random Test Generation: Random test era produces great outcome in test information era.It can haphazardly choose contribution until data sources are discovered [9]. Irregular test era is simple and also quick technique to produce test information. It can produce enormous number of experiments automatically. Symbolic Execution based procedures:symbolic

execution is a method for test-information era proposed by James King in 1976 [10]. Typical execution based system contain dynamic typical execution and concolic testing.Typical test information era methods [11,12] dispense typical qualities to the factors and create mathematical expressions for the numerous imperatives in the program code. An imperative solver is utilized to decide answer for these expressions that cover a test prerequisite.Typical execution can be utilized for various purposes, as discovery of bug, program check, troubleshooting of code, support, and limitation of blame [13].

In inquiry based techniques,search calculation name itself speaks to some sort of scope.It has extensive applications in test information era in light of the fact that the creating programming tests is an undesirable issue [14,15]. It use as enhancement method for test information era and to take care of programming building issues. Advantage of utilizing SBST is that its result demonstrates the productivity of approach just disadvantage is it requires extensive inquiry space.

A. Major issues involved in object oriented Testing

The issues of testing article arranged frameworks are legacy, embodiment and polymorphism. Alternate issues ofquestion situated frameworks are decentralized code,encapsulation,testcases tinguishing proof and raising special cases.Question arranged frameworks are worked out of at least two interrelated items A Determining the accuracy of O-O frameworks requires testing the techniques that change or convey the state ofan protest Å Testing strategies in a protest situated framework is like trying subprograms in process-situated frameworks .The conditions happening in traditional systems are information conditions between factors calling conditions between modules ,practical conditions between a module and the factors it processes, denitional conditions between a variable and its sort. OO Systems have extra conditions: class to class conditions class to strategy conditions; class to message dependencies,class to variable conditions strategy to variable conditions. technique to message conditions and strategy to strategy dependencies.[8]

II. LITERATURE SURVEY

Many Researches have proposed a wide range of solutions to reduce the number of test cases in object oriented system.

a.Soft Computing approaches

Bipin Pandey, Rituraj[1] Jain connected delicate Computing in various sorts of programming testing.By utilizing the systems for programming testing,the result has been upgraded and the general execution of testing was made strides

b.Testing least dependent class

Dr. Reena Dadhich and Sourabh Sehgal [2] favored the test arrange from minimum ward class to most ward class.The outcome demonstrated that most ward class to minimum ward class spares the time and endeavors.

c.Random test case generation

Ingle S.E.1, Mahamune M. R[4] clarified the various types of procedures for producing experiments to satisfy experiment examination standard. The way arranged test perceives way for which experiment must be made, still the way may be found infeasible, and the test information generator improves the info that is go through the way.

d.UML sequence diagram

A.V.K. Shanthi1 and G. Mohan Kumar[7] proposed the experiment era by methods for UML Sequence outline utilizing Genetic Algorithm where test cases are upgraded. This approach is essential to recognize area of a blame in the execution, in this way decreasing testing exertion. Besides this strategy for experiment era proposes the engineers to enhance the outline quality, lessen programming improvement time and discover blames in the execution early phase of programming development[5].

III. PROPOSED SYSTEM

In our proposed framework, we bring the protest under two noteworthy classification.1)Dependent question 2)independent question. We group the question utilizing reflexive technique.We give most elevated need to ward protest while doing programming testing.Because,dependent question covers more number of information members.Then,second need moves to free questions. We test both items utilizing JUnit tool. After ordering the question arranged testing, we proposed another strategy for decreasing the quantity of experiments without trading off the nature of the product. The foundation of our proposed strategy is reflection and JUnit tool.In this area we talk about the quality of reflection,issues in question arranged testing which conveys an ideal answer for protest situated testing. we talk about the parts of our approach one by one..

A.Reflection

Reflection is one of advanced features of object oriented languages. This technique has the ability to find the type of objects at run time. It supports various methods which helpful to find out whether an object is a dependent object or independent object. The proposed system uses the reflection technique to classify whether the object is an a independent object or dependent object. This can be used by testers and developers who have a strong grasp of the fundamentals of the object oriented language. since it is a sensitive technology, it requires a run time permission of the security manager.

B.Testing the object using JUnit tool

One of the cost effective software testing is JUnit tool. Many Researchers have utilized the effectiveness of JUnit tool to test the object based software testing. In our proposed approach we use rule based fuzzy logic[8]. To test all kinds of data types of data members of an object, we choose only one data member from each cluster.

IV. PROPOSED FRAMEWORK
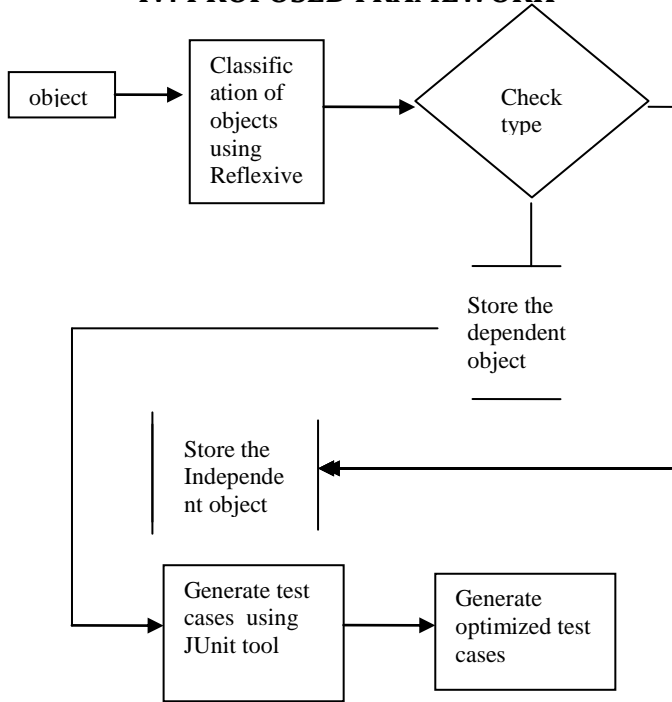
V. EXPERIMENTAL SETUP

Our experimental setup consists of an application which is implemented in java bean.These applications are executed in the Netbeans IDE.We utilized the power of java for classification of objects.The following Table1 shows the implementation code for classification of objects

```java
import java.io.*;
class sample{
public void method()
{
System.out.println("hai");
}}
class simple
{
public void meth(){
System.out.println("hoi");
}}
class sumple extends simple
{
public void met(){
System.out.println("welcome");
}}
class MainClass{
public static void main(String ar[])throws IOException
{
System.out.println("Enter the choice");

DataInputStream dis=new DataInputStream(System.in);

int x=Integer.parseInt(dis.readLine());

String name=" ";

Class c;

if(x==1)
{sumple s=new sumple();
c=s.getClass();
```


Then the dependent and independent objects are tested using JUnit tool distinctly.The following table shows tested the independent object using JUnit tool

```java
public class independent {
    public int add(String s1,String s2)
  {
    int int1=Integer.parseInt(s1);
    int int2=Integer.parseInt(s2);
return int1 +int2;
   }
  public int mul(String s1,String s2)
  {
    int int1=Integer.parseInt(s1);
```

The following table shows the dependent objects are tested using JUnit tool.In this each child class create a distinct test program due to challenging issues of object oriented software.In each class tested distinctly giving the input values.

```
public class dependent {
    public int add(String s1,String s2)
  {
    int int1=Integer.parseInt(s1);
    int int2=Integer.parseInt(s2);
return int1 +int2;
  } public int mul(String s1,String s2)
  {
    int int1=Integer.parseInt(s1);
    int int2=Integer.parseInt(s2);
return int1*int2; } }
class child extends dependent
    {
  public int sub(String s1,String s2)
  {
    int int1=Integer.parseInt(s1);
    int int2=Integer.parseInt(s2);
return int1-int2; } }
public class heri
```

VI. CONCLUSION

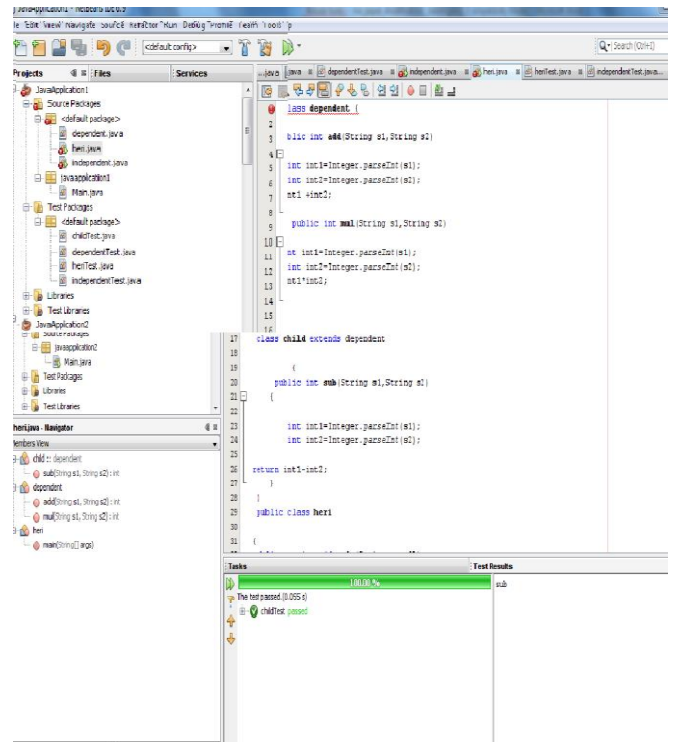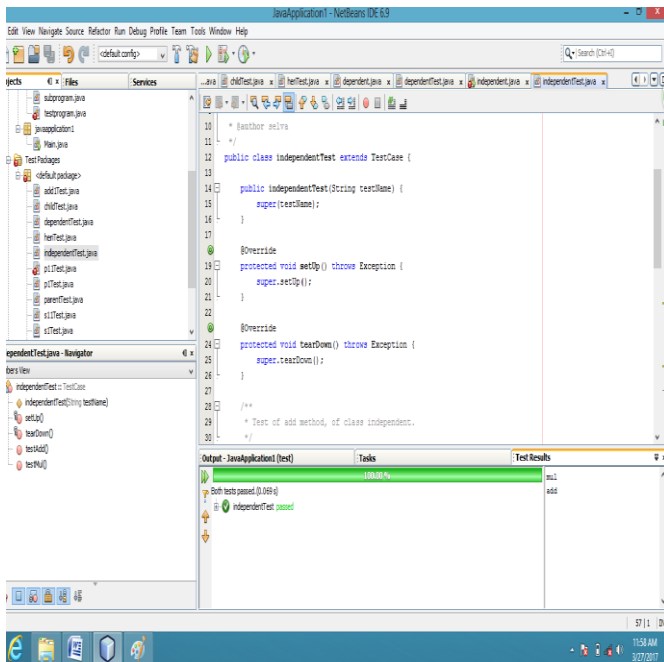In this paper, we exhibited an approach,that is order the protest utilizing reflexive procedure and store it.Then these dependent objects and independent objects are tested using JUnit tool.The dependent object and independent objects are tested using items are assembled utilizing bunching strategy. Bunching procedure gives an approach to choose a test information thing from each gathering and diminishes the quantity of testcases. Using sequence of method calls testing the dependent objects save the time and reduce cost framework resources.The extent of our proposed framework is completely mechanized apparatus to be developed.The device will encourages all levels of question arranged programming testing and spare time and decrease the cost of the product.

VII. REFERENCES

[1]. Bipin Pandey, Rituraj Jain, Soft Computing Based Approaches for Software Testing: A Survey, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4, Issue-2, May 2014.

[2]. Dr. Reena Dadhich1 and Sourabh Sehgal2 Analyzing The Efficient Test Order For Integration Testing,Advanced Computing: An

International Journal (ACIJ), Vol.3, No.4, July 2012(dependent independent object)

[3]. http://mit.bme.hu/~micskeiz/pages/code_based_test_generation.html (lastly accessed on April 05, 2016).

[4]. Ingle S. E.1, Mahamune M. R., An UML Based software Automatic Test Case Generation: Survey International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 02 | May-2015(Irjet-v2i276.pdf)

[5]. GurkarandeshKaur, Parminder Singh,Test Case Generation Using UML Diagram, International Journal of Emerging Technologies in Engineering Research (IJETER) Volume 1, Issue 2, July (2015)(Manuscripts-volume-1-issue)

[6]. A.V.K. Shanthi and G. Mohan Kumar, Automated Test Cases Generation from UML Sequence Diagram, 2012 International Conference on Software and Computer Applications (ICSCA 2012)IPCSITvol. 41 (2012) © (2012) IACSIT Press, Singapore(testcasesequences2.pdf)

[7]. Object-Oriented Software Testing | Some Research and Development David C. Kung and Pei Hsia

[8]. Computer Science and Engineering Dept. The Univ. of Texas at Arlington Yasufumi Toyoshima, Cris Chen, Jerry Gao Fujitsu Network Communication Systems

[9]. Jon Edvardsson ,"A Survey on Automatic Test data generation", 2nd Conference on Computer Science and Engineering , Vol. 2, No. 1, (October 1999) , pp. 21-28.

[10]. J. C. King, "Symbolic execution and program testing", Communication ACM, vol. 19, no. 7, (July 1976), pp. 385–394.

[11]. Howden, William E.,"Symbolic testing and the DISSECT symbolic evaluation system", IEEE Transactions on Software Engineering, vol. 3, no. 4, (July 1977), pp. 266-278.

[12]. John Clarke, Mark Harman, Bryan Jones, "The Application of Metaheuristic Search techniques to Problems in Software Engineering", IEEE Computer Society Press Vol. 42, No. 1, (2000), pp. 247-254.

[13]. L. A. Clarke and D. J. Richardson," Applications of symbolic evaluation", Journal of Systems and Software, (Feb 1985), pp. 15–35.

[14]. Tao Feng, Kasturi Bidarkar, " Survey of Software Testing Methodology", vol. 25, no-3 , (2008) , pp. 216-226.

[15]. Voas,Morell and Miller, "Predicting where faults can hide from testing", IEEE vol. 8 , pp. 41-48.

Achieving Security for Data Access Control Using Cryptography Techniques

[1]Dr.V.Vasanthi, [2]S. Akram Saeed Aglan Alhammadi, [3]Ramkumar. S, [3]Sathish Kumar

[1]Asst.Prof, Dept. of Computer Science,
[2]Ph.D Research Scholar, Dept. of Computer Science
Rathinam College of Arts and Science, Rathinam Techzone, Coimbatore, Tamil Nadu, India
E-mail :vasanthi.cs@rathinamcollege.com
[3]Asst.Prof, Dept. of Computer Applications,
[3]Asst.Prof, Dept. of CS & IT
Kalasalingam University, Madurai, Tamil Nadu, India

ABSTRACT

The amount of data being collected and stored every day by private and public sectors increased dramatically. Almost all industries, organizations and hospitals are maintaining personal information about individuals for decision making or pattern recognition. Security risk is very high while sharing this personal sensitive information among different data collectors. Therefore, privacy-preserving processes have already been developed to sanitize confidential information beginning with the samples while keeping their utility. For that safe and secure distributed computation new privacy preserving data mining algorithm has been developed. The main goal of these algorithms is to prevent that sensible information from hackers, during knowledge extraction from voluminous data. This work presents a protection saving approach that could be connected to decision tree learning, without associative misfortune of precision. This approach changeover the definitive specimen information sets into a gathering of undiscovered information sets, from which definitive information examines can't be remade without the whole assembly of unbelievable information sets. In the mean time, a proficient and precise decision tree might be manufactured straightforwardly from those stunning information sets. This novel methodology might be connected straightforwardly to the information space when the first sample is gathered. The methodology is versatile with other protection preserving approaches, for example, cryptography for extra protection.

Keywords : RSA, Data Mining, DSA, Cryptography, Cloud

I. INTRODUCTION

Data mining is an emerging field which connects different major areas similar databases, artificial intelligence and statistics. Data mining is an efficient tool for investigate and extract previously unknown patterns from voluminous data. The procedure of data mining requires a large amount of information to become collected into a central site. In modern days organizations are extremely depending on data mining in results to provide better resources, to achieving greater profit, and better decision-making. For these purposes organizations collect huge amount of data [1]. Just for example, business organizations collect data concerning the consumers for marketing purposes and improving

business strategies, medical organizations collect medical records for better treatment and medical research. When using the rapid ahead of time web, networking, hardware and software technology there is remarkable growth in the quantity of data that could be collected from several sites or organizations. Huge volumes of Data collected in this particular manner also have sensitive data about individuals. It has been obvious that whenever a data mining algorithm is run against the union of various databases, the extracted knowledge simply not only consists of discovered patterns and correlations that might be hidden within the data however it reveals the information which is regarded as to be private. Privacy is a vital issue a lot of data mining applications that cope with health care,

security, financial along with other types of sensitive [2][3].

Privacy preserving knowledge mining is fundamentally new research area, which effectively extracts hidden information without including private knowledge of individuals. Privacy preserving technique becomes the new direction for organizations sharing the knowledge for clustering and also achieving privacy of individuals. Privacy issues are thinking about in situations. They are centralized surroundings and distributed surroundings. In centralized surroundings, database is available in single location. In this surroundings, a privacy preserving knowledge mining techniques are used to hide sensitive knowledge of individuals [4]. In distributed surroundings, knowledge is distributed to multiple sites.

The objective of the computational task is participates the parties to securely process some functions of their distributed and private inputs. An inquiry that emerges here is the thing that it implies that for a computation to be secured. A system of approaching this address is to provide a list of properties that ought to be saved. The headmost such property that regularly comes to mind is that of confidentiality or privacy.

A naive attempt at formalizing privacy would be to require that each party learns nothing about the other parties' inputs, though if it behaves maliciously. However, such a definition is generally unattainable because the defined output of the computation typically reveals some information on other parties' inputs. (For example, a decision tree estimated on two distributed databases reveals some information about both databases.) Therefore, the privacy requirement is generally formalized by saying that the only information learned by the parties in the computation (again, even by those who behave maliciously) is that specified by the function output. Even though privacy is a primary security property, it seldom suffices.

Another important property is that of correctness; this states that the parties' output is really that defined by the function (if correctness is not warranted, then getting a malicious party probably can obtain the specified decision tree while the honest party receives a tree that is modified to offer misleading information). A central query that arises during this method of shaping security

properties is: when is our list of properties complete? This query is, of course, application-dependent and this essentially means that for each new drawback, the process of deciding which security properties are required must be re-evaluated. The effort that creating the right selection of properties will often be very challenging and it also may take number of years until we are believed strongly that a definition truly captures the security requirements that might be needed. Additionally, an incomplete of properties could simply contribute to real security failures. This paper concentrates on anticipating such attacks from third parties regarding the whole lifetime of the samples.

Contemporary research in privacy preserving knowledge mining chiefly falls into following categories first perturbation and randomization-based approaches, secondly secure multiparty computation (SMC)-based approaches. SMC approaches use cryptographic tools for collaborative information mining computation by multiple parties. Samples are distributed among different parties and they participate in the information computation in addition to communication technique.

SMC analysis focuses on protocol development for protecting privacy among the involved parties or computation efficiency; however, centralized processing of samples and storage privacy is out of the scope of SMC. A key utility of wide databases today is research, whether it is scientific or economic and merchandise oriented. Thus, for instance, the medical field has much to gain by pooling data for research; as can even competing businesses with mutual interests. Even though there is a potential gain, this is often unattainable because of the confidentiality issues which arise.

II. LITERATURE SURVEY

In recent years, privacy-preserving data mining techniques has been studied extensively, because of the wide proliferation of sensitive information on the internet. Most of the researchers and analyst focused on security and proposed number of algorithmic techniques for privacy-preserving data mining used in many applications such as medical, Bio-terrorism, identity theft, video surveillances, genomic and so on. In noise addition, generally a random number (noise) is drawn from a probability distribution having zero mean and a small standard deviation. This noise is then added to a

numerical attribute in order to mask its original value. Generally noise is added to the confidential attributes, of a micro data file before the data is released, in order to protect the sensitive information of an individual. However, adding noise to both confidential and Non-confidential attributes can improve the level of privacy by making re-identification of the records more challenging. The main objective of noise addition is to protect individual privacy by masking the micro data while introducing the least amount of incorrectness in it. The incorrectness in the statistic of a perturbed data set with respect to the statistic of the unperturbed data set.

Charu C. Aggarwal and Philip S. Yu [5] stated the methods which have been used in the privacy preservation Like: First, Randomization technique has been used to add the noise in the data in order to mask the attribute value of the records.

Hillol Kargupta ,souptik data,Qi wang and Krishnamoorthy sivakumar [6] triggering the randomized data distortion technique to mask the attribute value of the records for preserving the privacy of sensitive data. This work attempts to add the noise in the data values randomly and it's difficult to identify the predictive structure in spectral domain. To defeat this issue, work with the randomized object i.e. random matrix based spectral fitting techniques to retrieve the original values from the dataset and add random values. Xiaokui Xiao, Yufei Tao and Minghua Chen [7] achieves with two crucial properties. First, collusion is useless, meaning that the colluding recipients cannot learn anything more than what the most trustable recipient (among the colluding recipients) already knows alone. Second, the data each recipient receives and it can be regarded (and hence, analyzed in the same way) as the output of conventional uniform perturbation. The proposed technique is both spaces economical and computationally efficient.

Keke Chen and Ling Liu[8] created data mining service oriented framework to perform the geometric data perturbation approach and developed three protocols for perturbation unification for secure the geometric data with different parties. The unique feature of this proposed work has involves three protocol namely simply protocol, Negotiation protocol and space adaptation protocol. The basic challenge include (1) how to securely generate the same random perturbation in each site, while preventing the service provider knowing the unified perturbation, and (2) how to prevent privacy breach caused by data providers. Negotiation protocol overcomes the challenges of simply protocol by improving the overall privacy guarantee for all data providers. However, if perturbation data is generated randomly in simple protocol in terms of privacy guarantee then the providers may not be satisfied with in it execution. But in the negotiation protocol, each data provider has a chance to review the candidate perturbation and vote for or against the candidate

Alka Gangrade, Ravindra Patel[9] evaluated the multi party data to perform the privacy data perturbation approach with the secure manner and introduce a Privacy-preserving decision tree classifier using C4.5 algorithm without involving any third party. In multi party centralized the records are resides only in one party such horizontally or vertically to partitioning the data set. It is based on to calculate the union (frequent items can be evaluated using Secure Multiparty Computation methods for classifying rule mining) of all parties databases without trusted third party.

P.Kamakshi and A.Vinaya Babu[10] performed privacy preservation in the distributed environment using cryptography technique in data mining algorithms. Cryptography, the science of communication and computing in the presence of a malicious adversary extends from the traditional tasks of encryption and authentication to protocols for securely distributing computations among a group of mutually distrusting parties. In an ideal situation, in addition to the original parties there is also a third party called "trusted party" who does not deviate from the activities prescribed for him.

III. OVERVIEW OF THE WORK

The privacy preserving is one technique for changes over the original sample data sets into an aggregation of annoyed data sets. In that, original samples can't be recreated without the entire group of unreal data sets. In the meantime, an exact decision tree could be built directly from those unreal data sets.
This original approach can be applied directly to the data storage as soon as the first sample is collected. The approach is well-suited with other privacy preserving approaches, such as cryptography, for extra security.

The privacy is preserved even if the data is spread across multi parties. Suppose a bank issues the account holders' some of the attributes to more than one insurance agency. Then from the attributes of the table along with the records given to one insurance agency, other agency could not guess or identify the facts regarding the account holders. Likewise, if two agencies give their data set (retrieved from the bank) to other parties, they must not identify the facts by combing both data sets.
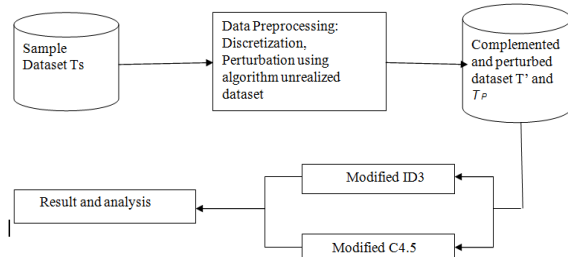
The overall frame work of the system involves two main modules first is data pre-processing and second is decision tree generation. In data pre-processing module initially continuous value attribute dataset is converted into discrete value after that the dataset is converted into sanitized version by using algorithm unrealized training set. Then generated complemented dataset and perturbed dataset is given as a input to decision tree generation module in which decision tree is built by using ID3 and C4.5 and result generated by both the algorithm is compared to analyze the algorithm.

IV. OVERVIEW OF THE TECHNICAL DETAILS

1 DEFINING PRIVACY PRESERVATION IN DATA MINING:

Basically there are two major dimensions in privacy preservation is: 1) Users' personal information 2) Information concerning their collective activity. First they have used individual privacy preservation and the latter as collective privacy preservation, which is related to corporate privacy. Individual privacy preservation: The primary goal of data privacy is the protection of personally identifiable information.

The information is considered personally recognizable if it can be connected, directly or indirectly, to an individual person. Thus, when personal information are subjected to mining, the attribute data associated with individuals are private and must be protected from disclosure. Analysts are able to acquire information from global models rather than particular individual characteristics.

2. COLLECTIVE PRIVACY PRESERVATION

It is not enough to protect only personal data. Sometimes, they may need to protect against learning sensitive knowledge representing the activities of a group. The protection of sensitive knowledge is collective privacy preservation. The goal is quite similar to statistical databases; within which security management mechanisms provide aggregate information about groups (population) and, at the same time, should prevent disclosure of confidential information about individuals.

However, unlike statistical databases, another aim of collective privacy preservation is to preserve strategic patterns that are paramount for strategic decisions, rather than minimizing the falsification of all statistics (e.g., bias and precision). In other words, the goal of collective privacy preservation is not only to protect personally identifiable information but also some patterns and trends that are not supposed to be discovered.

In collective privacy preservation, organizations have to deal with some interesting conflicts. For example, when personal information undergoes analysis processes that produce new facts about users' hobbies, shopping patterns, or preferences, these facts may be used in recommender systems to predict or affect their future shopping patterns. In general, this scenario is beneficial to both users and organizations. However, when organizations share data in a collaborative project, the goal is not only to protect personally identifiable information but also some strategic patterns.

In the business world, this pattern describes about the knowledge that can provide competitive advantages, and therefore must be protected. Protecting the knowledge that is discovered from confidential information is the big challenge (e.g., medical, financial, and crime information). The absence of privacy safeguards can evenly compromise individuals' privacy. While violation of individual privacy is clear, violation of collective privacy will lead to violation of individual's privacy.

Privacy Preserving Data Mining:

Models and Algorithms propose a number of techniques to perform the data mining tasks in a privacy-preserving way. These techniques generally fall into the following categories: data modification techniques, cryptographic methods and protocols for data sharing, statistical techniques for disclosure and inference control, query auditing methods, randomization and perturbation-based techniques. This edited volume also contains surveys by distinguished researchers in the privacy field. Each survey includes the key research content as well as future research directions of a particular topic in privacy.

3. Privacy Preserving Techniques

There are many approaches which have been adopted for privacy preserving data mining. To classify that based on the following dimensions:
- Data distribution
- Data modification
- Data mining algorithm
- Data or rule hiding
- Privacy preservation

The first dimension refers to the distribution of data. Some of the approaches have been developed for centralized data, while others refer to a distributed data scenario. Distributed data scenarios can also be classified as horizontal data distribution and vertical data distribution. Horizontal distribution refers to these cases where different database records reside in different places, while vertical data distribution, refers to the cases where all the values for different attributes reside in different places.

The second dimension refers to the data modification scheme. In general, data modification is used in order to modify the original values of a database that needs to be released to the public and in this way to ensure high privacy protection. It is important that a data modification technique should be in concert with the privacy policy adopted by an organization. Methods of modification include:
- ✓ Perturbation is accomplished by the alteration of an attribute value by a new value (i.e., changing a 1-value to a 0-value, or adding noise).
- ✓ Blocking is the replacement of an existing attribute value with a "?",

- ✓ Aggregation or merging which is the combination of several values into a coarser category,
- ✓ swapping that refers to interchanging values of individual records, and
- ✓ Sampling is refers to releasing data for only a sample of a population.

The third dimension refers to the data mining algorithm, for which the data modification is taking place. This is actually something that is not known before-hand, but it facilitates the analysis and design of the data hiding algorithm. We have included the problem of hiding data for a combination of data mining algorithms, into our future research agenda.

For the time being, various data mining algorithms have been considered in isolation of each other. Among that, the most important ideas have been developed for classifying data mining algorithms, like decision tree inducers, association rule mining algorithms, clustering algorithms, rough sets and Bayesian networks.

The fourth dimension refers to whether raw data or aggregated data should be hidden. The complexity for hiding aggregated data in the form of rules is of course higher, and for this reason, mostly heuristics have been developed. The lessening of the amount of public information causes the data miner to pro-duce weaker inference rules that will not allow the inference of confidential values. This process is also known as "rule confusion". The last dimension which is the most important refers to the privacy preservation technique used for the selective modification of the data. Selective modification is required in order to achieve higher utility for the modified data given that the privacy is not jeopardized. The techniques that have been applied for this reason are:

Heuristic-based techniques like adaptive modification that modifies only selected values that minimize the utility loss rather than all available values.
- Cryptography-based techniques like secure multi-party computation where a computation is se-cure if at the end of the computation no party knows anything except its own input and the results
- Reconstruction-based techniques where the original distribution of the data is reconstructed from the randomized data.

It is important to realize that data modification results in degradation of the database performance. In order to quantify the degradation of the data, its mainly use two metrics. The first one, measures the confidential data protection, while the second measures the loss of functionality.

V. Proposed Methodology

The existing ID3 (Iterative Dichotomiser 3) decision tree learning algorithm which covers the discrete-valued attributes that are implemented in the proposed system.

To preserve privacy when datasets are given to multiple parties, the proposed system finds the solution for the key problem of applying geometric data perturbation in multiparty collaborative mining which securely unify multiple geometric perturbations that are preferred by different parties, respectively.

In the proposed system, the privacy is preserved even if the data is spread across multi parties. Suppose a bank issues the account holders' some of the attributes to more than one insurance agency. Then from the attributes of the table along with the records given to one insurance agency, other agency could not guess or identify the facts regarding the account holders. Likewise, if two agencies give their data set (retrieved from the bank) to other parties, they must not identify the facts by combing both data sets.

The following are the advantages which are enclosed in the proposed theory

- ➢ Present the geometric perturbation approach which helps to multiparty privacy-preserving collaborative mining.
- ➢ From the two given data sets, the original facts can not be guessed.
- ➢ Privacy is preserved even if the data is spread across multi parties.
- ➢ Consider multiple service providers collaboratively providing the privacy preserving mining service to multiple data providers.

ID3 ALGORITHM

In decision tree learning, **ID3** (**Iterative Dichotomiser 3**) is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset. ID3 is the precursor to the C4.5 algorithm, and is often utilized in the machine learning and natural language processing domains.

Decision tree algorithms are a method for approximating discrete-valued target functions, during which the learned function is represented by a decision tree. These types of algorithms are famous in inductive learning and have been successfully applied to a broad range of tasks. Decision trees classify instances by sorting them down the tree from the root to some leaf node that provides the classification of the instances. Each node in the tree specifies some attribute instance and each branch descending from that node corresponds to one of the possible values for this attribute. The reasons for decision learning tree algorithms to be attractive are:
1. They generalize in a better way for unobserved instances, once examined the attribute value combined with in the training data.
2. They are efficient in computation as it is proportional to the number of training instances observed.
3. The tree elucidation gives an exceptional comprehension of instance and classify instances depend upon attributes
Example
ID3 algorithm is explained here using the classic 'Play Tennis' example.

VI. IMPLEMENTATION OF THE PROPOSED SYSTEM

7.1.1 ADD DATA SET

In this, the data set records are added. The outlook, temperature, humidity and windy column values along with result column is added with both realized and unrealized values. The details are saved into 'DataSet' table.

7.1.2 VIEW ORIGINAL DATA SET

In this, the data set records are viewed using data grid view control. The outlook, temperature, humidity and windy column values along with result column is displayed with realized values. The details are fetched from 'DataSet' table.

7.1.3 VIEW UNREALIZED DATA SET

In this, the data set records are viewed using data grid view control. The outlook2, temperature2, humidity2 and windy2 column values are added with unrealized values. The details are fetched from 'DataSet' table.

7.1.4 VIEW ORIGINAL/UNREALIZED DATA SET

In this, the data set records are viewed using data grid view control. The outlook, temperature, humidity and windy column values and outlook2, temperature2, humidity2 and windy2 column values are displayed. The details are fetched from 'Dataset' table.

7.1.5 MODIFY ORIGINAL/UNREALIZED DATA SET

In this, the data set records are viewed using data grid view control. The outlook, temperature, humidity and windy column values and outlook2, temperature2, humidity2 and windy2 column values are displayed. The details are fetched from 'DataSet' table. The data grid records can be modified and saved into database.

7.1.6 DECISION TREE GENERATION

In this, the decision tree is generated based on ID3 algorithm and displayed in the form. The records are taken from 'Dataset' table, best root attribute is selected and tree is constructed.

7.1.7 GEOMETRIC PERTURBATION

This module presents the geometric perturbation approach which helps to multiparty privacy-preserving collaborative mining. From the two given data sets, the original facts cannot be guessed. Privacy is preserved even if the data is spread across multi parties. It considers multiple service providers collaboratively providing the privacy preserving mining service to multiple data providers. The data is encrypted and extra bits are added in the result. In this approach, two results are added with two different bits. The application running in two places discards the unwanted bit and decrypts the data.

VII. Result and Discussion

The output accuracy (the similarity between the decision tree generated by the regular method and by the new approach), the storage complexity (the space required to store the unrealized samples based on the size of the original samples) and the privacy risk (the maximum, minimum, and average privacy loss if one unrealized data set is leaked).

ALGORITHM USED	ACCURACY	COST	EFFICENCY
Data Modification Techniques	LOW		MEDIUM
SMC Approaches	MEDIUM	HIGH	LOW
Perturbation-Based Approaches		HIGH	MEDIUM
Cryptographic Techniques			MEDIUM
Iterative Dichotomizer 3 Decision Tree Learning Algorithm		MEDIUM	HIGH
Multiparty Collaborative Mining	High	Low	High
Geometric Perturbation	High	Low	High

The privacy risk of the dummy attribute values technique, the average privacy loss per leaked unrealized data set is small, except for the even distribution case (in which the unrealized. Samples are the same as the originals). By doubling the sample domain, the average privacy loss for a single leaked data set is zero, as the unrealized samples are not linked to any information provider. The randomly picked tests show that the data set complementation approach eliminates the privacy risk for most cases and always improves privacy security significantly when dummy values are used.
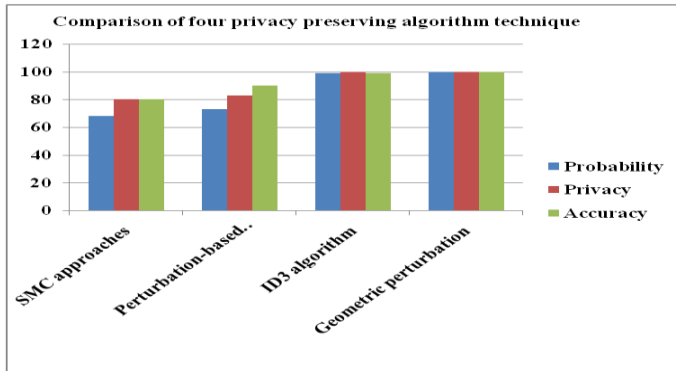
COMPARISON OF FOUR PRIVACY PRESERVING ALGORITHM TECHNIQUE

Algorithm	Probability	Privacy	Accuracy
SMC approaches	68	80	80

Perturbation-based approaches	**73**	**83**	**90**
ID3 algorithm	**99**	**100**	**99**
Geometric perturbation	**100**	**100**	**100**

Table 1 : comparison of four privacy preserving algorithm technique

Features	DES	RSA	Triple DES
Key Used	Same key is used for encryption and decryption purpose	Different keys are used for encryption and decryption purpose	Same key is used for encryption and decryption purpose
Scalability	It is scalable algorithm due to varying the key size and value size	No scalability occurs	It is scalable algorithm due to varying the key size and value size
Avalanche effect	No more affected	More affected	No more affected
Power Consumption	Low	High	More than DES
Throughput	Very high	Low	High
Confidentiality	High	Low	Very high

Figure 1 : Chart analysis of four privacy algorithm comparison

Table and fig 1 shows the comparison of four privacy preserving algorithm technique represents SMC approaches, Perturbation-based approaches, ID3 algorithm and Geometric perturbation. The result reveals that the SMC approaches given Probability, privacy and accuracy represents 68, 80 and 80. Perturbation-based approaches produced 73, 83, 90 and ID3 algorithm produced 99, 100, 99 finally Geometric perturbation technique given 100,100,100 we obtained the above results Geometric perturbation algorithm outperform the other remaining approaches.

THEORETICAL ANALYSIS

The theoretical analysis is as follow:

The security of the scalar product protocol is based on the inability of either side to solve k equations in more than k unknowns. Some of the unknowns are randomly chosen, and can safely be assumed as private. However, if enough data values are known to the other party, the equations can be solved to reveal all values. Therefore, the disclosure risk in this method is based on the number of data values that the other party might know from some external source.

The scalar product protocol is used once for every candidate item set. This could introduce extra equations. When the candidate item set contains multiple attributes from each side, there is no question of linear equations so it does not perceptibly weaken the privacy of the data.

Input Size (KB)	DES	3DES
75	21	57
96	32	55
112	54	81
286	97	173
359	188	198
600	198	202
951	391	327
5345.28	1399	1149
Throughput (MB/sec)	**3.01**	**2.8**

Table 2 : Execution Time (Milliseconds) of Encryption of different data packet size (DES & 3DES)
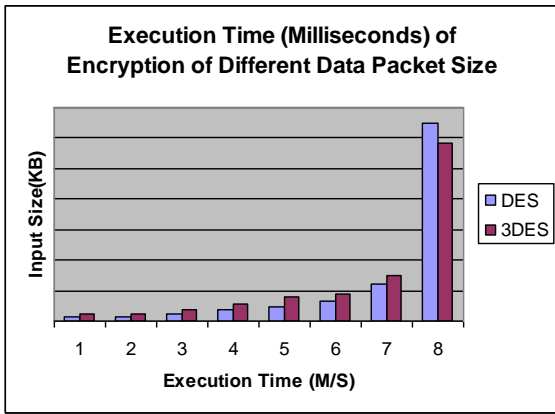
Figure 2: Execution Time (Milliseconds) Of Decryption Of Different Data Packet Size(DES And 3DES)

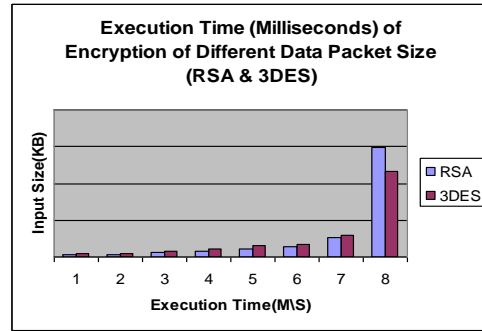| Input Size(KB)	RSA	3DES
45	42	45
55	37	42
96	58	65
236	88	104
319	152	135
560	148	160
899	252	181
5345.28	893	845
Throughput (MB/sec)	**2.0875**	**1.97125**

Table 2: Execution Time (Milliseconds) of Decryption of Different data packet size (DES & 3DES)

Input Size (KB)	RSA	3DES
75	38	50
96	35	44
112	55	76
286	83	113
359	105	155
600	143	177
951	264	299
5345.28	1296	1166
Throughput (MB/sec)	**2.52**	**2.08**

Table3: Execution Time (Milliseconds) of Encryption of Different data packet size
(RSA & 3DES)

Execution Time (Milliseconds) of Decryption of Different Data Packet Size(RSA and 3DES)

Input Size (KB)	DES	3DES
75	25	62
96	34	59
112	56	84
286	102	176
359	192	203
600	205	210
951	402	333
5345.28	1401	1152
Throughput (MB/sec)	**3.03**	**2.84**

Table 4: Execution Time (Milliseconds) of Decryption of Different data packet size (RSA & 3DES)
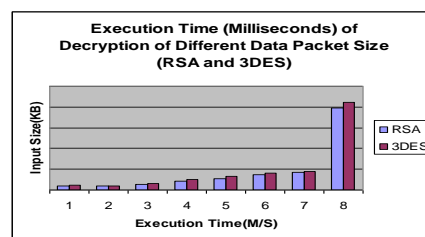
Execution Time (Milliseconds) of Encryption of Different Data Packet Size (DES and 3DES)

VIII. CONCLUSION

Privacy preservation via data set complementation fails if all training data sets are leaked because the data set reconstruction algorithm is generic. Therefore, further research is required to overcome this limitation. As it is

very straight forward to apply a cryptographic privacy preserving approach, such as the (anti)monotone framework, along with data set complementation, this direction for future research could correct the above limitation.

This paper covers the application of this new privacy preserving approach with the ID3 decision tree learning algorithm and discrete-valued attributes only. The norm in data collection processes, a sufficiently large number of sample data sets have been collected to achieve significant data mining results covering the whole research target. Second, the number of data sets leaked to potential attackers constitutes a small portion of the entire sample database.

This paper covers the applications of this new privacy preserving approach with the ID3 decision tree learning algorithm. In addition geometric perturbation mechanism is used so that the data is secured even distributed to more than one parties. It is suitable for multiparty data distribution.

IX. References

[1] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques" In Proc. of 3rd IEEE Int. Conf. on Data Mining, Washington, DC, USA,, pages99–106, 2003.

[2] R.Agarwal and R.Srikant, "Privacy preserving data mining", In Procseedings of the 19th ACM SIGMOD conference on Management of Data ,Dallas,Texas,USA, May2000

[3] J. Canny, "Collaborative filtering with privacy". In IEEE Symposium on security and privacy , pages 45-57 Oakland, May 2002.

[4] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques", In Proceedings of the 3rd IEEE International Conference on Data Mining, pages 99–106, Melbourne, Florida, November 19-22, 2003.

[5] C. Aggarwal and P. Yu, Privacy-Preserving Data Mining:, Models and Algorithms. Springer, 2008.restoneTireRecall.htm, May 2001.

[6] Hillol Kargupta ,souptik data,Qi wang and Krishnamoorthy sivakumar,"random data

Perturbation technique and privacy preserving data mining ",IEEE international conference on data mining, 2003.

[7] Xiaokui Xiao, Yufei Tao and Minghua Chen," Optimal Random Perturbation at Multiple Privacy Levels", ACM. VLDB '09, August 24-28, 2009, Lyon, France

[8] Keke Chen Ling Liu," Privacy-preserving Multiparty Collaborative Mining with Geometric Data Perturbation", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED COMPUTING, VOL. XX, NO. XX, JANUARY 2009

[9] Alka gangrade, ravindra patel," building privacy-preserving c4.5 Decision tree classifier on multiparties," international journal on computer science and engineering vol.1(3), 2009, 199-205

[10] p.kamakshi , dr.a.vinaya babu, " preserving privacy and sharing the data in distributed environment using cryptographic technique on perturbed data," journal of computing, volume 2, issue 4, april 2010, issn 2151‐9617

[11] Mohammad Ali Kadampur, Somayajulu D.V.L.N," A Noise Addition Scheme In Decision Tree For Privacy Preserving Data Mining," Journal Of Computing, Volume 2, Issue 1, January 2010, Issn 2151-9617

[12] Li Liu Murat Kantarcioglu and Bhavani Thuraisingham," Privacy Preserving Decision Tree Mining from Perturbed Data," Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009

[13] M. Naga lakshmi, k sandhya rani," a privacy preserving clustering method based on fuzzy approach and random rotation perturbation." Publications of problems & application in engineering research, vol 04, special issue01; 2013

[14] V.Thavavel and S.Sivakumar," A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment," International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012

[15] Pui k. Fong and jens h. Weber-jahnke," privacy preserving decision tree learning using unrealized data sets," ieee transactions on knowledge and data engineering, vol. 24, no. 2, february 2012

[16] S.nithya and p.senthil vadivu," efficient decision tree based privacy preserving Approach for unrealized data sets," international journal of advances in computer science and technology, 2(6), june 2013, 83 - 89

[17] Seema kedar, sneha dhawale, wankhade vaibhav," privacy preserving data mining," international journal of advanced research in computer and communication engineering vol. 2, issue 4, april 2013

[18] Justin Brickell and Vitaly Shmatikov," Privacy-Preserving Classi_er Learning",

[19] Keke chen ling liu," privacy-preserving multiparty collaborative mining with geometric data perturbation", ieee transactions on parallel and distributed computing, vol. Xx, no. Xx, january 2009

[20] S.L. Wang and A. Jafari, "Hiding Sensitive Predictive Association Rules," Proc. IEEE Int'l Conf. Systems, Man and Cybernetics, pp. 164-169, 2005.

[21] Q. Ma and P. Deng, "Secure Multi-Party Protocols for Privacy Preserving Data Mining," Proc. Third Int'l Conf. Wireless Algo-rithms, Systems, and Applications (WASA '08), pp. 526-537, 2008

[22] J. Gitanjali, J. Indumathi, N.C. Iyengar, and N. Sriman, "A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incre-mental Data Stream Privacy Preserving Data Mining," Proc. IEEE Second Int'l Advance Computing Conf. (IACC), pp. 410-415, 2010.

[23] L. Liu, M. Kantarcioglu, and B. Thuraisingham, "Privacy Preser-ving Decision Tree Mining from Perturbed Data," Proc. 42nd Hawaii Int'l Conf. System Sciences (HICSS '09), 2009.

[24] Y. Zhu, L. Huang, W. Yang, D. Li, Y. Luo, and F. Dong, "Three New Approaches to Privacy-Preserving Add to Multiply Protocol and Its Application," Proc. Second Int' l Works hop Knowledge Discovery and Data Mining, (WKDD '09), pp. 554-558, 2009.

[25] M. Shaneck and Y. Kim, "Efficient Cryptographic Primitives for Private Data Mining," Proc. 43rd Hawaii Int'l Conf. System Sciences (HICSS), pp. 1-9, 2010.

[26] J. Vaidya and C. Clifton. Privacy-preserving decision trees over vertically partitioned data. In Proceedings of the 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security, Storrs,Connecticut, 2005. Springer. L. Liu, M. Kantarcioglu, and B. Thuraisingham, "Privacy Preserving Decision Tree Mining from Perturbed Data," Proc. 42nd Hawaii Int'l Conf. System Sciences (HICSS '09), 2009.

[27] S. Ajmani, R. Morris, and B. Liskov, "A Trusted Third-Party Computation Service," Technical Report MIT-LCS-TR-847, MIT, 2001.

[28] Lian Liu.,Jie Wang.,Jun Zhang., ``Wavelet based data perturbation for simultaneous privacy preserving and statistics preserving.,'' In Proceedings of IEEE International Conference on Data Mining workshop., 2008.

A Smart Searching Technique for Optimizing Relevant Web portal Discovery

S. Dhanasekaran[1], Vamshikrishna Bandari[2], Ravi Teja[3], Vishnu Gupta[4]

*[1] Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India

srividhans@gmail.com[1]

*[2] Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India

krishnavamshi.12321@gmail.com[2]

*[3]Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India

teja.14raviteja@gmail.com[3]

*[4] Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India

thotavishnugupta@gmail.com[4]

ABSTRACT

This Research work is mainly deals about the Minimization of search options in a search engine. In general the keyword searched in any search engine gets some millions of results in microseconds. The output is obtained by analysing and processing a bulk data, thereby obtaining all relative or most searched sites and web pages. But here we put forth an idea for getting the fixed and efficient result in a short duration of time. The keyword searched in the search box gets the top 5 related sites per page through which the user can obtain the exact and efficient output. By this method of searching the search time for user and searching load for system and the server, both gets reduced which results in the effective usage of search engine.

Keywords: Search engine, crawling, spiders, building an index, API key, ranking factors.

I. INTRODUCTION

The frequently used search engine "Google" returns a bulk data even for a small search keyword typed. The returned data may or may not be required by the users totally. It causes great dissatisfaction if not required. So, this work introduced a new search engine "**Searchin** which is a modified version of Google. This new search engine gets the top most 5 results which are very much related and efficient to the users. There by reducing the search time. It is the modification of search engine API key.

II. METHODOLOGY

For optimizing or minimizing of search options we are going to use the search engine API key in our Research work. As we know that the best and fastest search engine in world is "GOOGLE", we need to modify the source code for optimization in a way that the top 5 results are displayed. For modification firstly we need to understand the search engine, that how it works, ranking factors.

Working of search in engine:
Search engines mainly perform two functions namely crawling and building an index. The term crawling is used because the search engine while

searching crawls each and every part of the web server and finds many billions of interconnected documents on the web. The interconnected documents are analyzed and best results are found. To complete these tasks search engines have constructed database centers all over the world. These database centers hold thousands of machines processing large quantity of data very quickly. Whenever we search anything in the search engine, we expect it to get the results as fast as possible, even a small delay causes great dissatisfaction. So, these engines work hard to obtain result as fast as possible. Whenever a keyword is searched the search engine searches in two major ways and obtain the result i.e.., the most visited and the most relevant, based on their previous searches and displays all in an order depending on their priority from first to last.
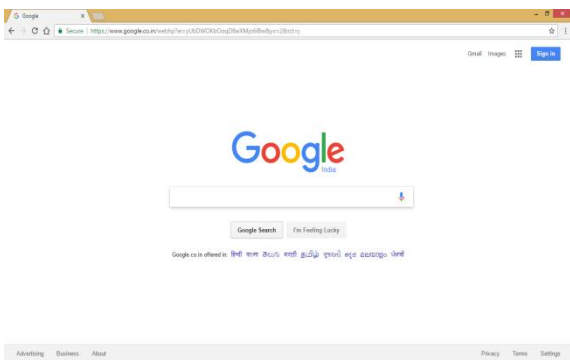
When a person performs an online search, the search engine find its billions of documents and does two things: first, it returns only those results that are relevant or useful to the searcher's query; second, it ranks those results according to the popularity of the websites serving the information. It is both relevance and popularity that the process of **SEO** is meant to influence. How do search engines determine relevance and popularity? To a search engine, relevance means more than finding a page with the right words. In the early days of the web, search engines didn't go much further than this simplistic step, and search results were of limited value. Over the years, smart engineers have devised better ways to match results to searchers' queries. Today, hundreds of factors influence

relevance, and we'll discuss the most important of these in this guide. search engines typically assume that the more popular a site, page, or document, the more valuable the information it contains must be. This assumption has proven fairly successful in terms of user satisfaction with search results. Popularity and relevance aren't determined manually. Instead, the engines employ mathematical algorithms to sort the relevance, and then to rank in order of quality (popularity). These algorithms often comprise hundreds of variables. In the search marketing field, we refer to them as "ranking factors."
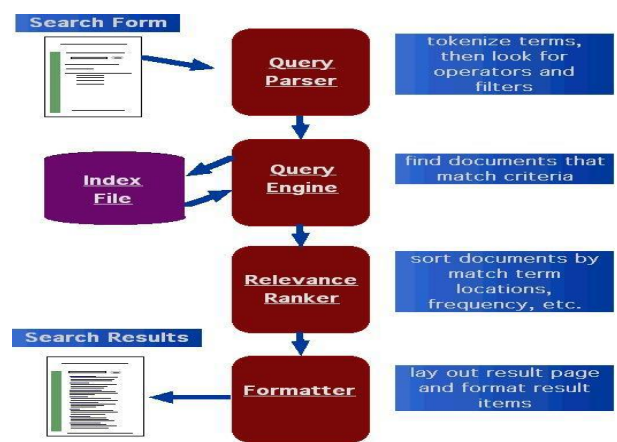
Ranking factors:

Ranking algorithm varied your search option in billion and trillions of pages in cloud to identify how similar to each one .Revealing the Companies own ranking algorithms is kept secret.

In past 3 years Google made a difficult to achieve it .earlier ,sites or web protocols are ranked on how much times a specific keyword /option was searched. These leads to "keyword stuffing ".where these pages or sites are most consists of meaningless matter as long as it provide the keyword or option in them.

Then the concept of importance based on linking was introduced more popular sites would be more linked to, obviously – but this led to a proliferation of spammed links all over the web. Now each link is determined to have a different value, depending

on the "authority" of the site in question. If a high-level government agency links to you, it's worth far more than a link found in a free-for-all "link directory".

III. EXPERIMENTAL RESULT

Ranking factor survey

Ranking factor Analysis

IV. CONCLUSION AND FUTURE SCOPE

On analyzing the working and ranking results of search engine we create a new search engine named as **"Searchin"** which function as Google because we used Google API key with slight modification as we want.

Smart searchin engine

The results come out by searching a keyword is based on ranking factors; cache stored in system and recently viewed pages. By this Searchin we can access data as fast as (exactly 10 times faster than **Google**). Based on the observation the following benefits has been achieved by our smart searching techniques Faster than Google in some aspect, Google filters are also accepted. Time consumption on searching is optimized compared to any other search engines, Results per page is five which allow the user to find best one. Data consumption is less, Algorithm is easy to design.

However this technique is going to find most relevant result with minimum time and data consumption. In future this searching techniques will be implemented in regional languages.

V. References

[1] Feng Zhao, Jingyu Zhou, Chang Nie, Heqing Huang, Hai Jin, "SmartCrawler: A Two- Stage Crawler forEfficiently Harvesting Deep-Web Interfaces", IEEE Transaction on services computing, Vol-99, 2015.

[2] M. Burner, "Crawling towards Eternity: Building an Archive of the World Wide Web," Web Techniques Magazine, vol. 2, pp. 37-40, 1997.

[3] Allan Heydon and Marc Najork, Mercator: A scalable,extensible webcrawler. World Wide Web Conference,2(4):219–229, April 1999.

[4]   Jenny Edwards, Kevin S. McCurley, and John A. Tomlin. An adaptive model for optimizing performance of an incremental web crawler. In Proceedings of the Tenth Conference on World Wide Web, pages 106–113, Hong Kong, May 2001. Elsevier Science.

[5]   Luciano Barbosa and Juliana Freire. Searching for hidden-web databases. In WebDB, pages 1– 6, 2005.

[6]   Soumen Chakrabarti, Martin Van den Berg, and Byron Dom. Focused crawling: a new approach to topic-specific web re-source discovery. Computer Networks, 31(11):1623–1640, 1999.

[7]   Jayant Madhavan, David Ko, Łucja Kot, Vignesh Ganapathy, Alex Rasmussen, and Alon Halevy. Google's deep web crawl. Proceedings of the VLDB Endowment, 1(2):1241–1252, 2008.

[8]   Olston Christopher and Najork Marc. Web crawling. Foundations and Trends in Information Retrieval, 4(3):175–246, 2010.

[9]   Balakrishnan Raju and Kambhampati Subbarao. Sourcerank:Relevance and trust assessment for deep web sources based on inter-source agreement. In Proceedings of the 20th international conference on World Wide Web, pages 227–236, 2011.

[10]   Denis Shestakov and Tapio Salakoski. On estimating the scale of national deep web. InDatabase and Expert Systems Applications, pages 780–789. Springer, 2007M.K.Sherbiny El "Efficient fuzzy logic load frequency controller"Energy Conversion and Management 43(2002)1853-1863 Elsevier

Author's Profile

DHANASEKARAN SUBBIAH working as Assistant Professor, at Kalasalingam University Srivilliputtur, Tamilnadu, India. His Native place is Srivilliputtur. He received the B.E degree in Computer Science and Engineering from Madurai Kamaraj University, Tamilnadu, India in 2004 and the M.E degree in Computer Science and Engineering from Annamalai University, Tamilnadu, India in 2007. He is pursuing Ph.D., in Kalasalingam University, Krishnankoil, and Tamilnadu, India under the Guidance of DR.V.VASUDEVAN, Eminent Professor and Registrar of Kalasalingam University. He has published many Research papers in Reputed International Journal and International conferences. He is a Life time member of ISTE (Indian Society for Technical Education).He is currently engaged in research in cloud computing and artificial intelligence.

Identification of Factors Affecting Governance of Data Security in Higher Education

Janet Gabriel[1], Dr.Hidayah Sulaiman[2]
[1]College of Computer Science & Information Technology, Universiti Tenaga Nasional, Malaysia
[2]Deputy Head of Department, Information Systems Department, College of Computer Science & Information Technology, Universiti Tenaga Nasional, Malaysia

ABSTRACT

Governance of data security in higher education is the data management mechanism that has to be practiced in all higher education. It is to have an appropriate management that employees in the organizations will understand the role to be played to secure confidential data. This paper aims at identifying the factors affecting governance of data security in higher education. It is to discover the issues in the management and the governance of data security in the higher education. Based on the literature review, new research model with eight factors were proposed for data security governance practices. Based on the research findings, a governance model was derived and proposed for higher education.

Keywords : Factors; Governance; Data Security; Higher Education

I. INTRODUCTION

Governance in the data security is about managing to protect the data in a secured manner. Management implements and makes specific decisions but involvement, accountability and decision rights are about security governance where both security and management are different [1]. Having responsibility, accountability and proper internal controls will protect informational assets in organizational structure by the information systems security governance [2]

Data governance states that proper security policies in information systems security governance are crucial. In the organization, faculty staffs need to have responsibility and accountability towards information security which is a sense of ownership [3].

The research has shown the governance of information security focuses in various fields. It is similar to the governance of data security. Every element concentrates on the security of data in whichever applications are being implemented and use by the entire user which has become a major importance. This is the main concern, in the sense that proper governance has to be established in

order not to have breach of data that will be violated without the knowledge of the user. This research will provide valuable governance to the faculty staffs in the higher education industry. It is to have a better understanding and be aware of the common factors that affect the governance in the area of data security. These data governance values have to be practice in the higher education to improve the governance and responsible for the directing and controlling issues of the organization [4].

This paper identifies the issues in management, factors affecting data security governance and propose data security governance model for better data security in higher education industry.

II. LITERATURE REVIEW

This paper focuses on the data security problems that occur in the higher education. Data security governance enhances tight security of governance in the higher education.

Importance of data security

Protecting information assets from being loss, damage, unauthorized access, operational breakdown, inaccessibility and misuse are the protection under the information security. Every data has important value compared to hardware device that data can be accessed by authorized user and unauthorized user at the same time which can cost high risk to security of the data [5].

Implementing a strategy for the information security in higher education that will meet the business objectives, needs and goals is imperative [4]. Important assets in the organization need to be protected by implementing the information security governance [6]. Institutions have to handle and give importance to the information security in order to secure information assets [7].

Benefits of data security

Data security starts from beginning of data transmission; until it will be safe and secured. Encrypting important data and granting access to authorized users to access data is one of the ways to protect the data [8] that will improve the operational effectiveness in the institution [4]. Hence security governance is partly based on the business processes in the organizations.

Challenges of data security

Access to related data must be strictly limited only to authorized users; otherwise, the privacy could be abused. Authentication problem can lead to invalid data [9]. New security criteria are needed to treat the threats on any system architectures that are affected [8]. The entire components of information system need to be secured, such as the data in the database management system. This is required when dealing with threats and outdated information security controls [10].

Challenges in the data security comes in the form of unauthorized access, faculty staffs may not be aware of data breach, mobile computing use and careless in misusing data [11]. Organization should provide training for all employees handling data, however financing the training is an issue in most of the organization [12].

Without the management support, accessing to resources legally or illegally can cause problems to the organizations [4]. Attacks on the information systems

and services can lead sabotage, crime, fraud, revenge, harassment and identity theft for the purpose of intellectual property or money. The desired place for the hackers to attack is the university environment which they can easily obtain confidential data [7].

Governance of data security in higher learning education

The governance in higher education includes formation of policies to monitor the work of the organization on a daily basis in the management or administration of the organization. Resources that need access requires conditions to control the access and key being generated, which permit the system to decrypt the resources.

Security and privacy of data are being risked when data are shared across system and data should be given assurance that it will be governed in a secured way with their policies [13] Various framework has been created to describe who can proceed, what actions with what information, when and under what conditions and what methods [8]. Development with the Information Technology (IT) Auditing and the mapping of CNSSI/NSTISSI standards with ISACA IT Auditing Model Curriculum, have assisted the public and private organizations for business systems and information technologies to be protected and controlled.

Each institution has different needs, resources and starting point. Anthony [14] pointed out that the main concern in the higher education is to protect the data, proper use, and the management of data as a business asset. It is organized on different models and goals than the corporate sector, non-profit organizations face similar issues such as risk, liability, business continuity, costs, and national repercussions when core activities were move to the Internet. Implementation of data governance tools and practices give benefits for the public higher education and provide quality of the institutional data, effect of the data governance tools and practices on institutional strategic planning and decision making [15].

Factors affecting governance of data security

There is a need to focus towards the data leakage, information security transformation, compliance

monitoring, cyber risks, cyber threats, disaster recovery, business continuity and prevention from data loss. There are six categories that require attention in the governance of the data security which is the security training, individual values, security policy implementation, beliefs, security culture and internal control assessment [3]. Most of the attacks comes from insider abuse, unauthorized access towards the systems, loss of data through portable devices for instance USB, PDA or phones [12].

Protecting and preserving the confidentiality, integrity, availability of information is the main concern of authenticity and reliability [11]. The occurrence and impact of information security breaches can be decreased through the learning and knowing of the computer security and ethics.

Data are to be protected through processes. The weakest connections towards the information systems are the faculty staffs in the organizations that they are to follow the information security rules and regulations [4]. Securing resources will be difficult if no management support and a security awareness program in the organization is restricted in what it is able to achieve [4]. Absence of information security process during recruitment by the human resource and training that is provided by human resource, variance demands from users, ignorance and attitude from users as well as inconsistently applied policies and procedures creates issues in the organizations when there is no management support [4]. Hence, communication becomes a problem when no proper communication tools and skills are available among the faculty staffs.

III. RESEARCH METHODOLOGY

This paper focuses on the identification of factors affecting governance of data security in higher education. The aim is to identify new data security governance practices were to help to solve data governance problems in the higher education.

The research carried out through quantitative approach in a form of collecting survey data. The data is analyzed by using statistical method. In the research design, qualitative method was used which consists of analytical approach focusing on the review and analysis of the literature review that will help to develop the research

conceptual model. A survey questionnaire is conducted to collect data and statistical tool is used to analyze the data. The research is based on the 8 hypotheses feedbacks and comments on the survey instrument. The data are collected from various faculties' staffs in a higher education institution. Then, the data is analyzed and to enhance the findings in the research. Furthermore, a qualitative approach is used for the data security governance model.

IV. FINDINGS AND DISCUSSION

This paper has been based on research conceptual model. Different governance data security models were proposed by various experts for different categories.

V. DATA SECURITY GOVERNANCE MODEL OF HIGHER EDUCATION

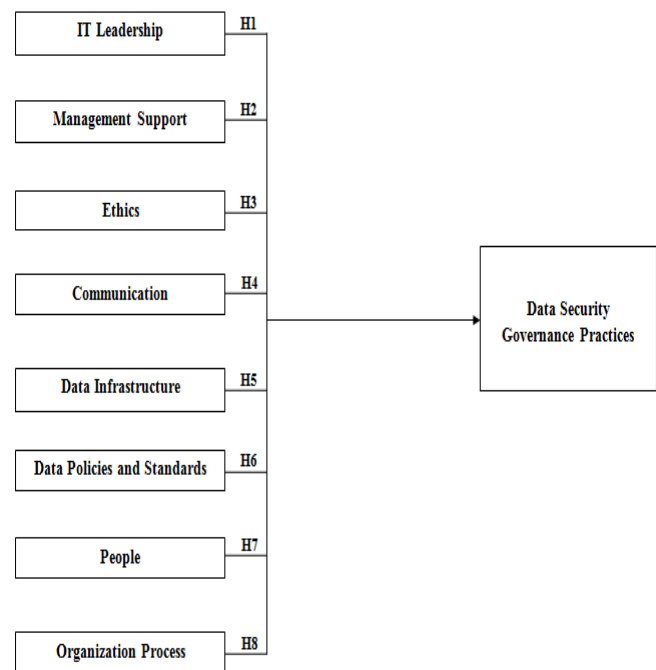Figure 1 illustrates the governance of data security conceptual model for higher education.

Figure 1: Governance of data security Model

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

s

Figure 2: pearson correlation formula

Table 1: anova (Stastical significance)

Pearson correlation is applied to identify the links between the variables. The objective of this research is to identify these factors that influence the data security governance practices in the higher education. It is to measure the linear correlation between two variables X and Y. The values of the Pearson Correlation range from -1 to +1 with negative numbers representing a negative correlation and positive numbers representing a positive correlation .The closer the value is to -1 or +1, the stronger the association is between the variables.
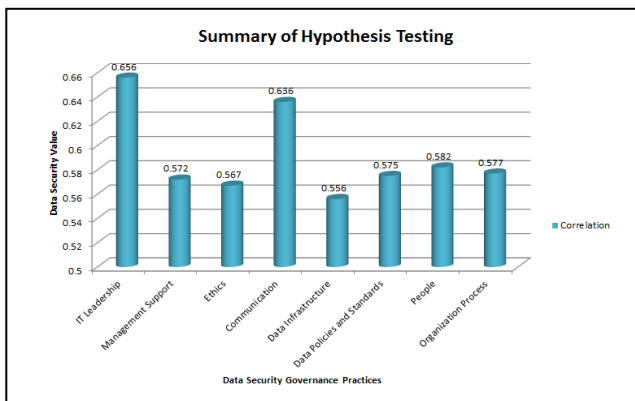
The F-ratio in the ANOVA table tests whether the overall regression is a good fit for the data. Table 1 shows that the independent variables statistically significantly predict the dependent variable, F(8, 106)=13.780,p<0.0005 which indicates that the regression model is a good fit of the data.

Model Summary				
Model	R	R Square	Adjustable R Square	Std. Error of the Estimate
1	.714ª	.510	.473	.54212
Predictors: (Constant), Organization Process, Data Policies and Standards, Management Support, Ethics, People, IT Leadership, Data Infrastructure, Communication				

Table 2: model summary of regression analysis

Figure 3: summary of hypothesis testing

The result generated from regression analysis is beneficial to show the impact of independent variables (IT Leadership, Management Support, Ethics, Communication, Data Infrastructure, Data Policies and Standards, People, Organization Process) in explaining the variation of the dependent variable (Data Security Governance Practices). Table 2 shows that the value of R Square is 0.510, which means the eight independent variables are able to explain 51% of the variation in the dependent variables.

$$Y = a + bX$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum x y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum x y) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Figure 4: Regression analysis formula

X is the independent variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0).
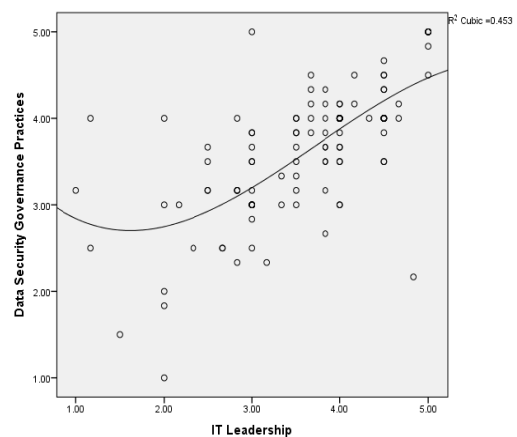
Figure 5: scatter dot graph

High positive correlation when the values increase together. 1 is a perfect positive correlation. 0 is no correlation and -1 is a perfect negative correlation. The Figure 5 shows that the values are more than 1 which indicate IT Leadership have positive relationship with Data Security Governance Practices.

		Sum of Squares	df	Mean Square	F	Sig.
Model						
1	Regression	32.397	8	4.050	13.780	.000ᵇ
	Residual	31.152	106	.294		
	Total	63.550	114			

ANOVAª

a. Dependent Variable: Data Security Governance Practices_Mean

b. Predictors: (Constant), Organization Process_Mean, Data Policies and Standard_Mean, Management Support_Mean, Ethics_Mean, People_Mean, IT Leadership_Mean, Data Infrastructure_Mean, Communications_Mean

VI. LIMITATION AND FUTUREWORK

There are some limitations that need to be taken into considerations through this research study. The respondents of this study are from various faculty staffs in the higher education. Their opinions may slightly differ among all the faculty members based on their exposure towards the governance of data security and expertise. This paper is adopting the quantitative approach using questionnaire to collect data. However, the difficulties to collect information from faculty staffs who are not aware on the importance of the governance in the data security may contribute to biasness in the results. Nevertheless, security measures in data management are indeed prioritized in the higher education. Future research is to empirically test the data security governance factors in confirming that these factors actually support the higher education's practices in data security governance.

VII.   CONCLUSION

This paper analyzes the governance of data security conceptual model for higher education in Malaysia. Based on the quantitative and qualitative survey, it was discovered that all the factors have positive relationship with the data security governance practices. As per summary of hypothesis, IT leadership is most influential factor that IT experts are expected to take lead control and organize the technologies in the higher education institution to have better data security governance in higher education. Research findings shows that the governance model of data security should include IT leadership, management support, ethics, communication, data infrastructure, data policies and standards, people and organization process.

VIII.   REFERENCES

[1]   Broadbent, M. (2003). Effective IT Governance by Design, Gartner Inc.

[2]   Dhillon, G. (2007) Principles of Information Systems Security: Text and Cases Wiley.

[3]   Sushma,M., & Gurpreet,D. (2016). Information systems security governance research: A behavioural perspective.

[4]   Munyaradzi.M., Patrick.M., and Theo.T. (2014).The Status of Information Security Governance within State Universities in Zimbabwe.

[5]   Sujithra, M., Padmavathi, G. & Sathya, N. (2015). Mobile device data security: A cryptographic approach by outsourcing mobile data to cloud.

[6]   Munirul.U.,Zuraini.b.I.,Zailani.M.S.   (2011).   A Framework for the Governance of Information Security in Banking System.

[7]   D.S.Bhilare.(2013). MOB-RAIL: A Framework for Implementation of Information Security Governance Process and Evaluation.

[8]   Rajesh,P., & Umesh,K,S. (2012). An overview and study of security issues and challenges in cloud computing.

[9]   Ming, L., & Wenjing, L. (2010). Data security and privacy in wireless body area networks.

[10]   Farrah, D.S.B., Norizan, A., Noraizan, A., & Riaza, P.M.R. (2014). Conceptualizing security measures on mobile learning for Malaysian higher education institutions.

[11]   Fazlida, M.R. & Jamaliah.S. (2015). Information security: Risk, governance and implementation setback.

[12]   Smith, M., Buchanan, W., Thuemmler, C., & Hazelhoff Roelfzema, N. (2010). Information governance and patient data protection within primary health care.

[13]   Knox, C.W., Jarl, N. & Steve, M. (2011) Persistent Security, privacy and governance for healthcare information.

[14]   Anthony. (2014). Data Governance Overview. Protiviti Risk Business Consulting.Internal Audit.

[15]   Marsha.P.(2011). Applying Aspects of Data Governance from the Private Sector to Public Higher Education.

Incremental-Parallel Data Stream Classification in Apache Spark Environment

A.Anantha Babu, J.Preethi

Department of Computer Science and Engineering, Anna University Regional Campus Coimbatore, Tamil Nadu, India

ABSTRACT

With notorious domain of big data age, the challenging task on data stream classification is high velocity conceptual infinite stream and perspective statistical properties of data which differs periodically. In this paper, we propose an Incremental Parallel Random Forest (IPRF) algorithm for data streams in spark cloud computing environment. The algorithm incrementally estimates the accuracy for classifying the data streams, which priors to parallelization process in order to reduce the training time and prediction process using random sampling and filtering approach, that improves the dynamic-data allocation and task-scheduling mechanism in a cloud environment. From the perspective of dynamic-data allocation, dynamically changes the data in a data stream environment, to reduce the communication cost, volume data using vertically data-partitioning, data-multiplexing method. From the perspective of task scheduling, an incremental-parallel technique is carried out in the training process of Random Forest and a task directed acyclic graph depends upon resilient distributed data objects as static, redundant, and least data object appending to re-organize the mapping relationship between successor task and slaves. The details and the results of evaluating the proposed mechanism using benchmark datasets are presented in this paper.

Keywords: Big Data, Data Stream Classification, Incremental-Parallel Technique

I. INTRODUCTION

A. *Motivation*

With the unprecedented growth and profuse appliance of cloud computing and Internet of Things (IoT), the transmission, sharing, and collection of data reached very high level. Data is produced and gathered continuously at a puzzling speed in such areas social networks, remote sensor network monitoring, on-line banking, and solar power system. Unlike traditional data sets, these emerging and popularizing data sets are enormous, temporally ordered, quickly changing, and potential with these characters are called as data stream [1].

Big data can be considered as one of the main source data streaming. The well-known challenging problems are data streaming likes Infinite length, concept-draft, and concept-evolution and feature evolution. One problem emerges when a data stream is assumed to

infinite length, yet it is speed and rapid phenomenon. Another issue is the concept draft presumption that streaming data can be underlying concepts of the stream change over time. A third assumption, when the concept and feature evolution can be used to predict the future behavior and also has been undermined as the technology continuously changed [2].

To refresh mining results are used to the promising approach in the incremental techniques of the high speed and analysis for large–scale data streams. Unfortunately, the new programming models are supported by the incremental processing. Data mining is processed to the hadoop famous cloud platform [3]. In the MapReduce model can be implemented to the machine learning algorithm, the Hadoop Distributed File System (HDFS) can be written in the intermediate results in all iterations. Enormous resource communication, storage and also disk I/O operation is taken much time and cost [4]. Data mining is suitable to be processed the cloud platform apache-spark. Apache spark is supported to the Resilient

Distributed Datasets (RDD) and Directed Acyclic Graph (DAG) in compared to hadoop. It can be performed to the stored data cache, perform computation and iterative computation for same data directly from memory. The slaves can be processed disk I/O operation to take the huge amount of time. Therefore, it can be more suitable for data mining with incremental computation [5].

L.Breiman is introduced to the Random Forest (RF) algorithm [6] at 2001. It can be supported the ensemble machine learning to data streaming classification, decision tree and bagging, weighted voting method to construct large number decision trees using probabilistically. Decision trees are constructed concurrently to be supported parallelization.

B. *Our Contributions*

In this paper, we propose an Incremental Parallel Random Forest (IPRF) algorithm for big data that is implemented on the spark cloud computing environment. The IPRF algorithm is optimized based on an incremental-parallel technique combining the data–parallel and task–parallel optimization. To improve the classification accuracy of IPRF, an optimization is proposed prior to the parallel process. Our contributions in this paper are summarized as follows.

- An incremental-parallel learning technique is proposed to improve the accuracy of IPRF, which combines the ideas of streaming decision trees and Random Forests.
- A dynamic duration data streams, in which the duration at each time step may dynamically change, this is in contrast to the traditional data–parallel optimization where the duration is fixed and static.
- Based on the dynamic data allocation, a memozation task parallel optimization is proposed and implemented on spark. Training task DAG of the data stream is constructed based on the RDD model and Memozation task schedulers are involved in perform the tasks in the DAG.

This paper we divide into different section: the section I is the introduction. Section II reviews the related work that has been done towards data stream mining and gives the limitations of the conventional methods and fundamental techniques for constructing dynamic data allocation and task scheduling. In Section III, we introduce two fundamental techniques for constructing

our Random Forest in data streaming classification: Filter method and weighted voting method. Section IV describes data stream implementation of the RF algorithm on spark. Section V presents the experimental results and evolutions, with respect to the classification accuracy and performance. Finally, Section VI presents the conclusion and future work.

II. RELATED WORK

Although traditional data processing techniques have achieved good performance for static data, they are difficult to be handled to dynamic data efficiently. When a dataset becomes the data arrive at high speed and in huge volumes, the deemed concept of a data stream is frequently subject change along the time and large size, the accuracy and performance data mining algorithm are significantly denied [7, 8].

Due to the need of address the underlying concepts change over a time and noisy data, various improvement method introduced by researchers. Hulten et al. [9] and Bifet et al. [10] introduced some data streaming classification algorithm for mining time–changing data streams to address the issue of window size. These algorithms use Concept–Adapting Very Fast Decision Tree (CVFDT) and Hoeffding Adaptive Tree using Adaptive Windowing (HAT–ADWIN) and achieve efficient variable–length window in binary classification problems. Wang et al. [11] proposed a general framework for classification of data streams with concept–drift. Kanoun et al. [12] and Ghazikhani et al. [13] focused on the defect of adaptive ensemble algorithm is that the basic classifier learning model must have the ability to learn incrementally for new data blocks. Tarkoma et al. [14] proposed a theory and practice of bloom filters for distributed systems. Based on the existing research results, we propose a new optimization approach in this paper to address the problem of infinite length, concept–draft, concept–evolution, feature evolution and noisy data, which reduce the training time of the data according to the structure of the RF and improves the algorithm's accuracy with a low computational cost.

Focusing on the performance of classification algorithms for streaming data, numerous studies on the intersection of parallel / distributed computing and the learning of decision tree model were proposed. Palit et al. [15]

proposed a parallelized boosting with Map-Reduce, in which parallel ADABOOST and LOGITBOOST algorithms that achieve parallelization in both time and space. Svore et al. [16] carried out a boosted decision tree ranking algorithm, which addresses the speed and memory constraints by distributed computing systems.

Focusing on dynamic resource allocation and task scheduling execution in a parallel and distributed environment, Warneke et al. [17] proposed a dynamic resource allocation for efficient parallel data processing in a cloud environment. Chen et al. [18] carried out locality-aware and energy-aware job pre-assignment for MapReduce. Liu et al. [19] proposed a dynamic resource allocation and task scheduling strategy with uncertain task runtime on IaaS clouds. Zhang et al. [20] proposed an evolutionary scheduling of dynamic multitasking workloads for big data analysis in an elastic cloud.

Apache Kafka provides a distributed server for message queuing and then it works on the underlying concept of publish–subscribe mechanism. To provide large scalability for storing and consuming data are various machines in the cluster to be distributed on the streaming data. Due to replicated and persistent storage, it curbs up the risk of data loss [21]. Spark streaming is an extension of the core Spark API that enables scalable, high-throughput, fault-tolerant stream processing of live data streams. Spark is programmable framework for massive distributed processing of datasets, receives input data streams, divides the data into batches and then processed by the spark engine to generate the results.

III. Dynamic Random Forest Algorithm Optimization

We proposed an optimization approach for the RF algorithm to the improvement of the reduce time and space, classification accuracy for data streams. First, a pre-processing is performed in the classification. Second, a base classifier is model constructed in the pre-processed data. After ensemble approach is used to the output of multiple algorithms combined with the streaming data.

A. **Pre–Processing: Sampling and Filtering for Streaming data**

To improve the accuracy of the RF algorithm for streaming data, we present a new pre-processing method to reduce the time and memory the processing input data according to the importance of the feature variables. In a sampling techniques are constructing subsample from income streaming data according to some criterion. The stratified random sampling is used to create the sample of streaming data constructed to the models in ordered to subsample created into equal size. Filtering or feature selection is one of the common processes on the data streams. Filter methods can be performed rank individual feature or evaluate entire feature subsets. When the stream of data arrives, then the tuples which meet the specified condition will be accepted and others are dropped. The purpose of bloom filter in streaming analytics is the space and time efficient data structure. It deals with hash functions and makes the process of filtering more efficient. It can be solved by choosing an optimal number of hash functions and bit array size on the false probability. The process of filtering is presented in Fig. 1.
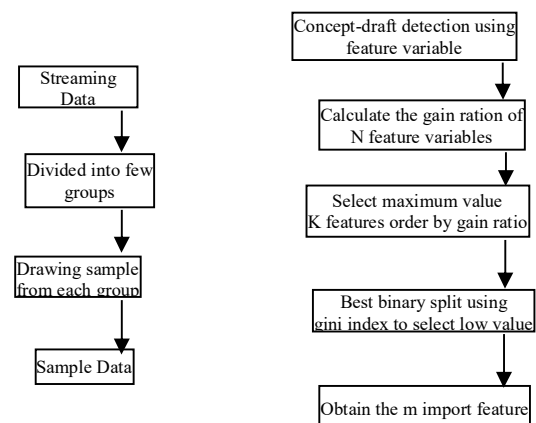

Figure 1. Stratified Random

Sampling and Filtering Process.

First, in the training process of each decision tree, Shannon's entropy is a measure of the indecorum associated with a random variable. It is defined as

$$H(y) = \sum_{i=1}^{n} p_i \log_2(p_i) \quad - (1)$$

Where y is a discrete random variable, and p_i is the probability of occurrence of y_i. To measure the entropy

for a dataset with known distribution requires using following equation.

$$H(y) = \sum_{y \to i = 1}^{n} P(y) \log_2 (P(y))$$ - (2)

Where P(y) is the probability mass function of y. A Two–window paradigm is used entropy in the context of detecting concept changes in data streams.

Second, the self-split information I(y$_{ij}$) of each input variable is calculated, as defined

$$I(y_i) = \sum_{k=1}^{o} - p_{(k,i)} \log_2 (p(k,i))$$ - (3)

Where o is the number of different values y$_i$ and P(k,i) is the probability of the type of value k with all type variables in y$_i$.

Then, the information of each feature variable calculated,

$$IG(y_i) = H(y) - H(y_i)$$ - (4).

By using information gain to measure the attribute, with the maximum gain ratio selected as the splitting attribute, it can be avoid the over fitting problem. As the split information approaches 0, the ratio becomes unstable. The gain ratio is defined as

$$GR(y_i) = \frac{IG(y_i)}{I(y_i)}$$ - (5)

Third, Gini index to select the best attribute, to measure how well separate on the feature variable. Decision tree is growing at each level to measure the best split and resume operation continue on the best split.

$$GI(y) = 1 - \sum_{i=1}^{n} P_i^2$$ - (6)

In a binary split compute a results on the partitioning to calculate on the weighted sum of the impurity. It can be split on y partitions y1 and y2 on the gini index of y that partitioning is

$$Gini(y) = \frac{y_1}{y} Gini(y_1) + \frac{y_2}{y} Gini(y_2)$$ - (7)

For each attribute possible to binary split is considered. The splitting subset can select on the minimum gini index for that attribute. The splitting attribute selected on the reduction of impurity to maximized on the attribute. In a splitting criteria can be considered on the attribute, splitting point and splitting subset.

$$Gini(A) = Gini(y) - Gini_A(y)$$ - (8)

Fourth, the hoeffding bound is used to accumulate an enough records in a robust decision for attribute test to be made. The hoeffding bound states that, given a random variable s in the range L, m independent values of s having mean (s) and the true mean of r is at least (s)-e, where upon

$$E = \sqrt{\frac{L^2 \ln(1/g)}{2m}}$$ - (9)

With probability 1-g and g is a user-defined threshold.

B. Random Forest on Streaming data classification

Random forest ensemble process is constructs on the basis of two techniques stratified random sampling and bloom filtering. The pre-processed data requires less amount of space and time when compared to the processing input data. Decision trees store the data in the leaf node and it is a balanced tree which is often linked to one another. Decision tree makes incrementally updated on the evolution of concept draft. Random forest is used for classification which consists of multiple decision trees. The output of each instance is calculated based on the averaging or weighted voting method the predictions of trees using ensemble techniques.

First, a pre-processing is performed in the classification and it can be handled to the huge and high–velocity data. Data streaming scenarios are considered to the most important parameters (Memory and Time). Pre-processed data can produce more or less equal to that of accuracy obtained through original data on the pre-processing techniques. To reduce the training and testing

time is used to the stratified random sampling and bloom filtering.

Second, after pre-processed data is constructed to the classification on the decision tree. Data streaming is considered to the concept–draft, feature selection, changes over a time, and model stability. Classification process can be performed to the perfectly predicts the unseen data, we can say there is a model stability. The classification process is constructed to the multiple decision tree model.

Third, Ensemble approach has been used for classification and it has been used to the supervised machine learning models. Machine learning algorithm has been trained and predictions to the unseen data than the models build using single classifier. Ensemble represents a single hypothesis and it seems that ensemble models have more flexibility.

C. **The Standard Random Forest Algorithm**

L.Breiman has developed the random forest algorithm for ensemble classification techniques based on the decision tree model. A number of binary decision tree grows as with any tree ensemble classifier and each new record can predict the class using a plurality of the class predictions from the set of trees. However, it can differ from standard ensemble techniques in the way to grow each tree are selected in which records and the way at each internal node are selected in which attributes.

Suppose that a data set contains p records, each with q attributes. The steps for each tree growing of the random forest algorithm are as follows:

Step 1: Choosing k different subsets of the p records from a training set in a stratified random sampling manner.

Step 2: For each internal node, a subset of q (q<<p) stratified randomly chosen attributes are selected, and the way in which attribute and the split point is a decision made using the standard Gini index algorithm only the selected attribute.
Step 3: The tree is left pruned.

The stratified random selection of records are replacement leave off some records, are never used in the building this tree. To estimate the error rate of each tree can be used these records. This process can be repeated with a new subset of the records to produce a user specified number of trees. Each tree is made the choice of the attribute and split point at each internal node. This tree is being built to depend on the other records and other attributes that were chosen to build this internal node.

IV. Incremental-Parallel learning of the Random Forest algorithm on Spark

We propose Incremental-parallel learning of the Random Forest algorithm on apache spark, to address the problems on the real-time streaming data communication cost and workload imbalance problem of large scale data in a parallel and distributed environment. The IPRF algorithm is optimized based on the dynamic data allocation and task scheduling for dynamic parallel optimization. The dynamic data allocation optimization on the vertical data-partitioning and data multiplexing method are performed each iteration results to summarized. The filter method reduced the data size and improve communication cost on the distributed and parallel environments. In a task scheduling based on data allocation to run the task carried out in the training process of the IPRF and RDD objects based on the DAG task created. Then task schedulers are minimize the data communication cost.

1. **Dynamic data communication on spark cluster**

Apache spark is a fast and general engine for large-scale data processing in a data streams with unbounded sequence. The apache spark can be perform two primitives: a Master and a Slave. It can be performed in a Resilient Distributed Datasets (RDDs) collection objects to compute a results. RDDs can be contains two actions: Transformations (convert into new form on existing data) and Operations (computing results). It can be supported SQL operation and can be avoid on the redundancy execution using Directed Acyclic Graph (DAG). The spark is an advantages vertically partitioned data, data multiplexing method and dynamic data allocation on a streaming environment. Then it can be achieved into workload balance, reduce communication cost achieved on the spark.

The vertical data partitioning method is a gini index computing task of each feature variable take up dynamic of the training time. In a task can be changed dynamically based feature variables. IPRF can performed reduced volume of data and distributed cost in a parallel and distributed environment. The training datasets are divided into dynamic feature subsets.

The training datasets are vertical partitioned into subsets by using feature subsets. There is a subset construct to the individual RDD objects and independent of other subsets. There is a dynamic feature variable added on end of the RDDs objects. The volume of datasets are sampling problems arise on the linearly increment on the RF scale and can be created Data-Sampling-Index (DSI) on the table. In data multiplexing methods are perform: allocate to slaves on the dynamic sampling times, allocate feature subsets to the clusters and generate DSI table by using gini index computing on the dynamic nature of datasets. Gini index is performed into dynamic feature variable selection on the training samples.
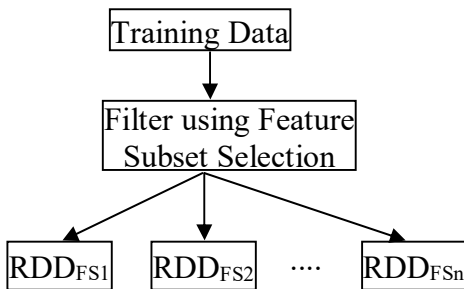

Figure 3. Vertical Data Partitioning Method

i. **Dynamic workload and task allocation strategies**

To better dynamic workload allocation strategies on data-partitioning and data-multiplexing method is performed on the feature subsets. In a dynamic nature of datasets are varied on each time based on the data type and volume. The datasets can be allocated into multiple slave nodes in a volume data. In a data allocation method have a three types: greater than slave capacity, less than slave capacity and equal amount to slave capacity. There is a data allocation method into local and global communication of data in the cluster environments. The RDD objects are allocated to the particular slave node and after execution results can be stored on persistent ways. The dynamically objects are changed to the feature subsets and same object allocation to the same worker node to avoid the execution.

The decision trees are built using IPRF on the dynamic independent feature variables construct the sub-node. There is some gini index computing task change on dynamic feature selection. It can be performed: training process and testing process. In a training process are trained to the filtering and sampling into dynamic data to be a feature selection, and testing process are performed into DAG create to construction on the process execution independent slaves. In a compute task scheduling can divided into two types: gini-index splitting task and node-splitting tasks. Gini-index is a computing task split into multiple slave nodes and impurity levels are reduced on the construction of the decision tree. Node splitting process are task executed on the worker nodes and momzation is stored results on the execution of the cluster worker nodes.
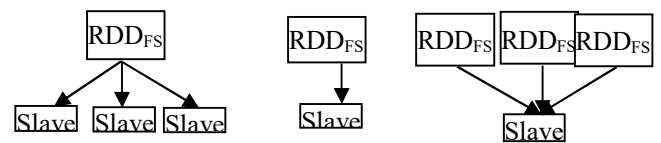

Figure 4. Dynamic Data Allocation and Task execution strategies

V. Extensive Experimental Results

A. Experimental Setup

The apache spark cloud platform performed all experiments and it is built of one master and three slave nodes. Each node contains the Linux Mint 18, Intel Pentium inside 2GHz GPU and 8GB memory. All the nodes are connected at high-speed Gigabit network and are configured with hadoop 2.7.3, Spark 2.1.0 and the algorithm is implemented in scala 2.11.0. The streaming datasets with time and memory is used in the experiments.

We have used bench mark data name Solar Power System which consists 1,88,835 instances with 40 attributes which contains 5 quantitative variables, 4

binary wildness, 6 solar power system attributes and no missing attributes. In a solar power system datasets used to detect the income amount energy or power, the usage of energy or power and amount of energy or power saved in the feature usage. In this three types of operation can be executed on spark cloud platform. In a training data not load into spark cluster. The spark cluster can be processed the datasets results to be stored. Then each iteration can execute the single process on the RDD objects process the slave node and result can be stored to the master node. There is a next iterations data to check before process the data, then conditions are

1. the incoming data is new to be process
2. some modification on the existing data to process
3. Existing data can't processed then already processed results to retrieve in the master.

B. Classification accuracy

The classification accuracy is not affected on the model construction using samples and filtered tuples. To evaluate the actual, sample and filter data has been implemented on the random forest using spark. The spark can be converted into libSVM format (label index: value index: value) for pre-processing. The algorithm is mainly considered to the some parameters: fixed (number of class) and dynamically changing parameter (maximum number of bins, maximum depth of tree and maximum number of trees). The majority voting methods are used to prediction on test data. The streaming data is mainly considered to the factors on time and memory. Then time taken to test the incoming data for model construction.

The test error is calculate using confusion matrix contains the information on the actual and prediction classifications done by random forest ensemble method. Then a binary prediction can be considered:

- X and U be number of correct and incorrect predictions on the instance of class a.
- Y and V be number of correct and incorrect predictions on the instance of class b.

$$Errorrate = D+U \,/\, X+D+U+Y \qquad - (10)$$

The error rate can be identified to varies size of data execute on the spark environment and then the results are represented to the Fig. 4.

Error rate

DTS (Data Streaming) Size
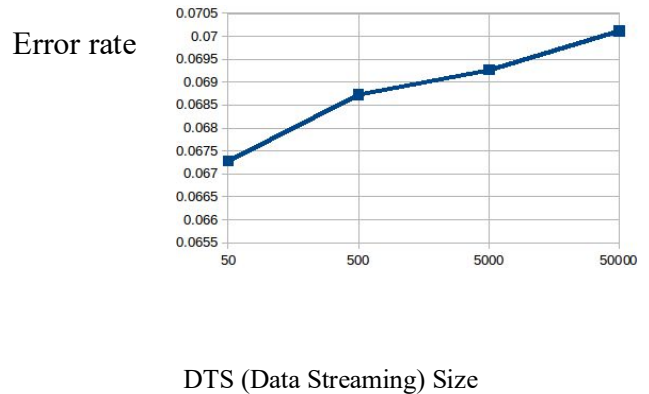
Figure 5. Test accuracy for different streaming data **size**

Testing Time (Seconds)
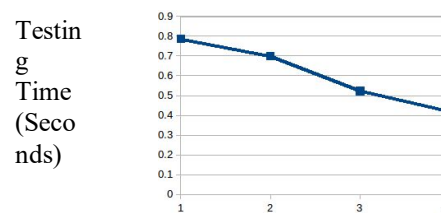
Figure 6. Testing Time analysis

The random forest is execute the data on the test and model time varies for actual, sample, filtered, both into represents Fig. 5. The curves are represented on the model building time and test on the instance. The model build and testing time can be reduced on the actual data execution process to sampling and filtering process data.

Model Building time (Seconds)

Sample Data
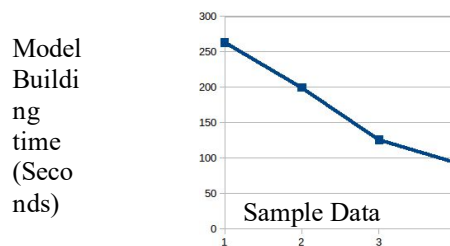
Figure 7. Model Construction analysis

C. Spark Cluster Scaling

The system generate amount of data is based on the available data reported by master on each slave node, the dynamic data optimization problem, can be solved using linear programing methods. Then, the slave requests

master data based in the optimization problems solutions. If the slave request is accept or reject on the master node to be process the data. Then some tasks are not assigned to the successfully, the master assign the task to large on small number slave processed.

Figure 8. Dynamic Data allocation and Task

D. **Data Communication Cost Analysis**

In the data communication cost can be analyzed on the scaling on the spark cluster in the IPRF algorithm. In the master node monitor the data size, then shuffled to the data among slave nodes. In this case, data size reduced ways to communicate the slave nodes in the cluster. Here, data is not transfered full size, the specified condition data of user request data to only.
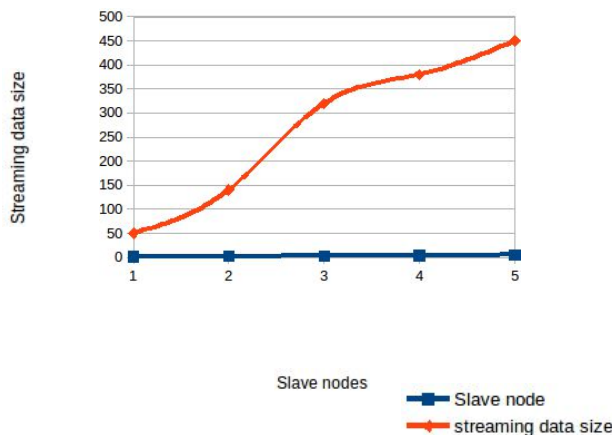

Figure 9. Data Communication Cost analysis of IPRF

VI. Conclusion

In this paper, an incremental-parallel random forest algorithms has been implemented for data streams. The classification accuracy of the IPRF algorithm is optimized through sampling and flitering method using weighted-voting method. Then, a incremental-parallel technique combining dynamic data allocation and task scheduling optimization is performed and implemented in a spark cloud computing environment. The feature subset selection have designed a framework for application oriented dynamic resource allocation, task scheduling and scalable data allocation algorithm to minimize the cluster deployment cost constrained on capacity and service delay bound and the size of the reduce task with task duration.

VII.   Reference

[1] X. Wu, X. Zhu, and G.Q. Wu, "Data mining with big data," Knowledge and Data Engineering, IEEE Transactions on, vol. 26, no. 1, pp. 97-107, January 2014.

[2] M. M. Masud, Q. Chen, L. Khan, C. C. Aggarwal, J. Gao, J. Han, A. Srivastava, and N. C. Oza, "Classification and adaptive novel class detection of feature-evolving data streams," Knowledge and Data Engineering, IEEE Transactions on, vol. 25, no. 7, pp. 1484-1497, July 2013.

[3] Apache, "Hadoop," Website, January 2017, http://hadoop.apache.org.

[4] S. del Rio, V. Lopez, J.M. Benitez, and F. Herrera, "On the use of mapreduce for imbalanced big data using random forest," Information Sciences, vol. 285, pp. 112-137, November 2014.

[5] Apache, "Spark," Website, January 2017, http://spark-project.org

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, October 2001.

[7] P. Domingos and G. Hulten, "Mining high speed data streams," In: Proceedings of the 6th ACM SIGKDD international conference on knowledge discovery and data mining (SIGKDD), August 2000, pp 71–80.

[8] A. Tsymbal, "The problem of concept drift: definitions and related work," Technical report TCD-CS-2004-15, Computer Science Department, Trinity College Dublin, Ireland.

[9] G. Hulten, L. Spencer, and P. Domings, "Mining time-changing data streams," In: Proceedings of the 7th ACM SIGKDD international conference on knowledge discovery and data mining (SIGKDD), August 2001, pp 97–106.

[10] A. Bifet and R. Gavalda, "Adaptive Parameter-free learning from Evolving Data Streams,"

Technical report, Polytechnic University of Catalonia, 2009.

[11] H. Wang, W. Fan, P. Yu and J. Han, "Mining concept-drifting data streams using ensemble classifiers," In: 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), August 2003, pp. 226-235.

[12] K. Kanoun and M. van der Schaar, "Big-data streaming applications scheduling with online learning and CDF detection," In: Proceedings of the 2015 Design, Automation and Test in Europe Conference and Exhibition, EDA Consortium, March 2015, pp. 1547-1550.

[13] A. Ghazikhani, R. Monsefi, and H.S. Yazdi, "Online neural network model for non-stationary and imbalanced data stream classification," International Journal of Machine Learning and Cybernetics 5, no. 1, pp. 51-62, February 2014.

[14] S. Tarkoma, C.E. Rothenberg, and E. Lagerspetz, "Theory and practice of bloom filters for distributed systems," Communications Surveys Tutorials 14, IEEE, no. 1, pp.131–155, First 2012.

[15] I. Palit and C. K. Reddy, "Parallelized boosting with map-reduce," In: Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, pp. 1346-1353, December 2010.

[16] K. M. Svore and C. J. Burges, "Distributed stochastic aware random forest efficient data mining for big data," in Big data (BigData Congress), 2013 IEEE International Congress on, Cambridge University Press, 2013, pp. 425-426.

[17] D. Warneke and O. Kao, "Exploiting dynamic resource allocation for efficient parallel data processing in the cloud," IEEE transactions on parallel and distributed systems 22, no. 6, pp. 985-997, June 2011.

[18] L. Chen, J. Zhang, L. Cai, Z. Deng, T. He, and X. Wang, "Locality-Aware and Energy-Aware Job Pre-Assignment for Mapreduce," In: Intelligent Networking and Collaborative Systems (INCoS), 2016 International Conference on, pp. 59-65. IEEE, 2016.

[19] S. Liu, K. Ren, K. Deng, and J. Song, "A dynamic resource allocation and task scheduling strategy with uncertain task runtime on IaaS clouds," In: Information Science and Technology (ICIST), 2016 Sixth International Conference on, pp. 174-180. IEEE, 2016.

[20] F. Zhang, J. Cao, W. Tan, S. Khan, K. Li, and A. Zomaya, "Evolutionary scheduling of dynamic multitasking workloads for big-data analytics in elastic cloud," IEEE Transactions on Emerging Topics in Computing 2, no. 3 pp. 338-351, August 2014.

[21] Apache Kafka, A high-throughput distributed messaging system source: http://kafka.apache.org/ accessed July 2015.

An Information Based Dynamic Load Balancing Algorithms for Asymmetric Web Server Cluster through Secured Layered Networks

S. Tamilarasi[1], Dr. KungumaRaj[2]
[1]Department of Computer science,Research Scholar of ,Mother Teresa University, Kodaikanal, Tamilnadu, India
sstamilarasi@yahoo.com
[2]Department of Computer Applications,Head , Assitant Professor of Arulmigu Palaniandavar College for Women Palani, Tamilnadu, India

ABSTRACT

Networking is growing and changing perhaps even faster than other computer disciplines. Networks are both fragile and strong. The remote correspondence insurgency is conveying major changes to information systems administration, media transmission, and is making incorporated systems a reality. Load Balancing is the path towards circling workload transversely over various devices on a framework in handling. It intends to limit reaction time, amplify throughput, advance asset and stay away from over-burden of any single asset. To deal with secure substance over Web trade a protected channel given by Secure Socket Layer convention is proposed. Rather than customary system design that has a few weaknesses, we propose the outline and execution of Open Flow-based server bunches for dynamic load adjusting. Load balancing adjusting partitions activity between system interfaces on a system attachment (OSI display layer 4) premise, while channel holding infers a division of movement between physical interfaces at a lower level, either per bundle (OSI show Layer 3) or on an information connect (OSI demonstrate Layer 2) premise with a convention like most limited way crossing over.
Keywords: Layers, Load Balancing, Workload, Traffic

I. INTRODUCTION

Powerful load balancers intelligently determine which device within a given server farm is pleasant capable of manner an incoming statistics packet. Doing so requires algorithms programmed to distribute hundreds in a specific way.

Algorithms vary widely, depending on whether a load is sent on the network or application layer. algorithm choice affects the effectiveness of load distribution mechanisms and, therefore, performance and commercial enterprise continuity.

There are layered architecture in the network applications. The top most layers are such as Application Layer, Presentation layer not but not the least Network layer.

These layers are undoubtly involving in the architecture. With respect to these layers working behind the algorithms. Network layer and application layer algorithms differ in how they're able to analyze incoming traffic and the criteria they use to distribute traffic loads. Lets take the distribution logic. The network layer algorithms are Statistical or Randomized method can be used. On other hand the Application layer algorithm the Data Driven techniques are used.

In the case of Server load visibility option there is no permission in the network Layer algorithm. But on the application layer it will be visible to see the Server loading.

The primary challenge to network layer load balancers is a lack of visibility into site visitors glide, restrained to information stored in community packet headers. Routing decisions should be based totally on only a few elements—in the main source and vacation spot IP facts.

Community layer load balancers cannot verify the nature of incoming requests, their anticipated load technology and to be had server sources at a given time. A positive quantity of guess estimation is needed for them to make routing choices.
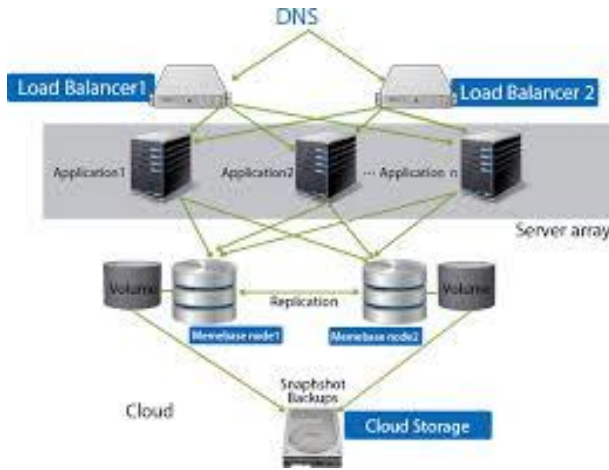
Examples of network layer algorithms include:

II. ALGORITHMS:

Round robin – A batch of servers are programmed to handle load in a rotating sequential manner. The algorithm assumes that each device is able to process the same number of requests and isn't able to account for active connections.

Weighted round robin – Servers are rated based on the relative amount of requests each is able to process. Those having higher capacities are sent more request

Least connections – Request are sent to the server having the least number of dynamic associations, accepting all

trifling if the application parts are stateless. Since the parts don't hold any value-based express, any of them can deal with similar demands similarly. On the off chance that all the more handling force is required, you simply include more servers and introduce the applications areas.
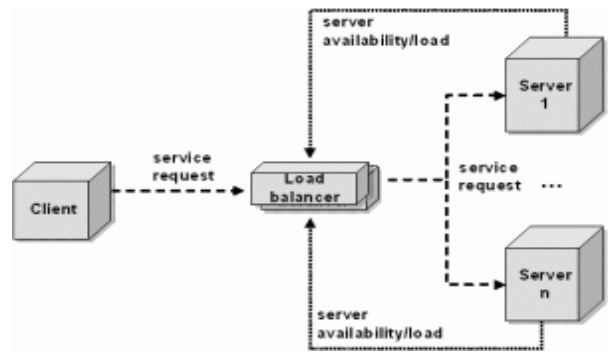
III. SERVER-LOAD BALANCING TECHNIQUES:

When all is said in done, server load balancing arrangements are of two primary sorts:

Transport-level: Load balancing -, for example, the DNS-based approach or TCP/IP-level load adjusting - acts autonomously of the application payload.

Application-level: stack adjusting utilizes the application payload to settle on load adjusting choices.

Load balancing arrangements can be further grouped into programming based load balancers and equipment based load balancers. Equipment based load balancers are particular equipment boxes that incorporate application-particular coordinated circuits (ASICs) modified for a specific utilize.
Network Load Balancing servers (also called *hosts*) in a cluster communicate among themselves to provide key benefits, including:

Adaptability: Arrange Load Balancing scales the execution of a server-based program, for example, a Web server, by disseminating its customer asks for over different servers inside the group. As activity expands, extra servers can be added to the group, with up to 32 servers conceivable in any one bunch.

High-accessibility : Organize Load Balancing gives high accessibility via consequently identifying the disappointment of a server and repartitioning customer movement among the rest of the servers inside ten seconds, while furnishing clients with persistent administration.

IV. LOAD BALANCER - FEATURES:

Awry load: A proportion can be physically allocated to bring about some backend servers to get a more prominent share of the workload than others. This is now and again utilized as an unrefined approach to represent a few servers having more limit than others and may not generally function as wanted.

Need actuation: When the quantity of accessible servers dips under a specific number, or load gets too high, standby servers can be brought on the web.

HTTP security: a few balancers can conceal HTTP mistake pages, expel server identification headers from HTTP reactions, and encode treats so that end clients can't control them.

Need lining: otherwise called rate molding, the capacity to give distinctive need to various movement.

V. MOTIVATION AND RESEARCH CHALLENGES

Distributed computing otherwise dispersing frameworks have turned out to be progressively prevalent as financially savvy modify local to conventional elite figuring stage . The principle point of load balancing issue on heterogeneous conveyed computational conditions is a productive mapping of undertakings to the arrangement of figuring hubs. The dynamic load balancing issue remains a testing worldwide enhancement issue due to the: (i) heterogeneous structure of the framework, (ii) processing assets managerial areas and (iii) Quality of Service(QoS) demands by applications.

The major motivation that leads to study
dynamic load balancing strategies in HDCS are listed as:

1. The figuring capacity of HDCS can be abused by outlining proficient errand designation calculations that sole out each assignment to the best reasonable processing hub for execution.

2. Due to heterogeneity of processing hubs, employments experience distinctive execution times on various registering hubs. Accordingly, research ought to address planning in heterogeneous conditions.

3. As dispersed frameworks keep on growing in scale, in heterogeneity, and in differing organizing innovation, they are introducing challenges that should be routed to meet the expanding requests of better execution and administrations for different circulated application.

4. Because of the recalcitrant way of the undertaking task issue on HDCS, it is attractive to acquire a most ideal arrangement through the outline of new systems for element stack adjusting in HDCS.

5. The assignments and figuring assets could be powerfully added and dropped to and from the framework. This requires dynamic load adjusting calculations that utilize framework state data for load task.

VI. CONCLUSION

Load balancing manages parceling a program into littler errands that can be executed simultaneously and mapping each of these undertakings to a computational asset, for example, a processor (e.g., in a multiprocessor framework) or a PC (e.g., in a PC organize). By creating systems that can outline assignments to processors in a way that equalizations out the heap, the aggregate preparing time will be decreased with enhanced processor usage. In any case, hereditary calculations have increased gigantic prominence in the course of the most recent couple of years as a vigorous and effortlessly versatile inquiry method. The work proposed here explores how a hereditary calculation can be utilized to take care of the dynamic load-adjusting issue.

A dynamic load-adjusting calculation is created whereby ideal or close ideal errand allotments can "advance" amid the operation of the parallel registering framework. The calculation considers other load-adjusting issues, for example, limit arrangements, data trade criteria, and entomb processor correspondence.

VII.   REFERENCE

[1]   X. K. Hwang, G.C. Fox, and JJ Dongarra. Distributed and Cloud Computing: From Parallel

Processing to the Internet of Things. Morgan Kaufmann, 2012. cited at p. 1, 2, 18]

[2] Hagit Attiya and Jennifer Welch. Distributed Computing: Fundamentals, simulations, and Advanced Topics. Wiley Series on Parallel and Distributed Computing.John Wiley and Sons Inc., 2000. cited at p. 1]

[3] T.V. Gopal, N.S. Nataraj, C. Ramamurthy, and V. Sankaranarayanan. Load balancing in heterogenous distributed systems. Microelectronics Reliability, 36(9):1279-1286, 1996. cited at p. 1, 9]

[4] H.J. Siegel, H.G. Dietz, and J.K. Antonio. Software support for heterogeneous computing. ACM Computing Surveys (CSUR), 28(1):237-239, 1996. cited at p. 1]

[5] Jie Wu. Distributed System Design. CRC press, 1999. cited at p. 1, 3, 5, 6, 19, 25, 101]

[6] A.Y. Zomaya and Y.H. Teh. Observations on using genetic algorithms for dynamic load-balancing. Parallel and Distributed Systems, IEEE Transactions on, 12(9):899-911, 2001. cited at p. 1, 4, 5, 6, 8, 24, 25, 42, 69, 70, 75, 76, 85, 86, 89, 93, 99]

[7] M. Maheswaran and H.J. Siegel. A dynamic matching and scheduling algorithm for heterogeneous computing systems. In Heterogeneous Computing Workshop, 1998.(HCW 98) Proceedings. 1998 Seventh, pages 57-69. IEEE, 1998. cited at p. 2]

[8] M. Maheswaran, S. Ali, HJ Siegal, D. Hensgen, and R.F. Freund. Dynamic match- ing and scheduling of a class of independent tasks onto heterogeneous computing systems. In Heterogeneous Computing Workshop, 1999.(HCW'99) Proceedings. Eighth, pages 30-44. IEEE, 1999. cited at p. 2, 30]

[9] M. Maheswaran, T.D. Braun, and H.J. Siegel. Heterogeneous distributed computing. Wiley Encyclopedia of Electrical and Electronics Engineering, 1999. cited at p. 2t p. 2]

[10] Jean Dollimore George Coulouris and Tim Kindberg. Distributed Operating System-Concepts and Design. Addison Wesley, second edition, 2000. cited at p. 2]

[11] Vijay K. Garg. Elements of Distributed Computing. Wiley-Interscience: JohnWiley and Sons, Inc. Publication, 2006. cited at p.2]

[12] Sukumar Ghosh. Distributed systems: an algorithmic approach. CRC press, 2010. cited at p. 2, 3]

[13] Andrew S. Tanenbaum and Maarten van Steen. Distributed Systems: Principles and Paradigms. Pearson Education, Inc., 2002. cited at p. 2]

Secure Socket Layer Based Load Balancing Methodology In Distributed Servers

[1] G. Srilakshmi,[2] Dr. K. Kungumaraj

[1]Department of Computer Science, Mother Teresa Women's University, Kodaikanal, TamilNadu, India
gsrilakshmi.gsl@gmail.com[1]
[2]Department of Computer Applications, Arulmigu Palaniandavar Arts and Science College for Women, Palani,TamilNadu,India
kungumarajkalimuthu@gmail.com [2]

ABSTRACT

A SSL load balancer goes about as the server-side SSL endpoint for associations with customers, implying that it plays out the decoding of solicitations and encryption of reactions that the web or application server would some way or another need to do. The procedure differs fairly relying upon the security of the system between the heap balancer and server. On the off chance that the heap balancer and server are on the same secured organize (for the most part this implies being behind a firewall), the SSL load balancer is typically arranged to unscramble the demand, separate the data required for load adjusting, and forward the demand to the server free (decoded). It encodes the server's reaction before returning it to the customer. On the off chance that the system between the load balancer and server is not secure, the SSL load balancer is generally arranged to decode the demand, remove the data required for load adjusting, and re-encode the demand before sending it to the server. The procedure is turned around for the reaction from server to customer. Offloading the decoding and encryption handle, which is computationally serious, liberates web and application servers to play out the work they are intended for, which speeds content conveyance and enhances the general client encounter. In the event that the system between load balancer and servers is secure, you just need to introduce and deal with the SSL declarations on the load balancer rather than each web and application server. This fundamentally diminishes managerial overhead if the gathering of servers is expansive.

Keywords: Load balancing, SSL, SSL Methodology, Load balancing Techniques

I. INTRODUCTION

In a cluster based server pool or network server, all solicitations from customers to an application server are initially passed to a wholesaler from a Web switch and after that the merchant advances each demand to one of the application servers as indicated by its circulation approach. The circulation in the application server ought to be done another way contrasted with the front-level Web server in which a store mindful conveyance like Locality-Aware Request Distribution indicates great execution. Particularly because of the high overhead of the SSL convention, the wholesaler in an application server ought to embrace an arrangement that limits the SSL overhead. Since the session reuse conspire, which is broadly utilized as a part of single Web servers, is exceptionally viable to diminish the SSL overhead, the

specialist plan to misuse the session reuse plot for the group based application servers.

The distributor of SSL-Session calculation keeps up the customer data to forward resulting demands from a similar customer to a similar application server. The upside of the SSL-Session is that it stays away from the pointless verification and arrangement stage when a customer tries to reconnect to the server. The weakness of this model is that it might bring about load awkwardness among the servers. On the off chance that a server has customers whose solicitations are visit and require much element calculation, the solicitations can't be disseminated to other gently stacked servers, bringing about load skewness among the servers. The fundamental disadvantage in the current technique is time idleness. The server load couldn't be adjusted

because of capricious heap of the intermediary servers. The current framework neglects to query way length and the quantity of overwhelming hubs experienced in every way. The metric of way length mirrors the execution of the question sending plan, and the metric of number of overwhelming hubs indicates how the blockage control convention maintains a strategic distance from substantial hubs in direct activity stream.

The system developed using Round Robin and SSL-Session model are not effective. Those models are not able to give the output in time and the throughput also lesser than that the expected output. Round Robin and SSL-Session models had the latency problem and minimal throughput. For this reason, the researcher introduced the SSL-LB load balancing model. The following are the drawbacks in the existing system:

1. In the existing system, users need to spend more time for retrieving their data.
2. Time consuming
3. Latency Problem
4. Minimum Throughput

II. SSL-LB METHODOLOGY

The SSL-LB model is aimed at mitigating the limitation of SSL-Session by using a new load balancing mechanism to achieve load balancing. The SSL-LB load balancing module get the load information from each application server and navigate the client request to the lightly loaded application server.

For effective load balancing, this examination presents Secure Socket Layer plot, which is utilized to associate the customer with the intermediary servers. Proficient diverting can be made by the server when a client asks for at the server top time. SSL-LB strategy will decrease the idleness time and increment the throughput than the current framework (Round Robin display and SSL-Session). The SSL-LB will lessen the server pay stack and enhances the throughput while the server is being occupied.

SSL-LB, which operates between the HTTP and Transmission Control Protocol (TCP) network layers, is the most popular tool that provides a secure channel between a client and a Web server. SSL-LB is composed of two components: a handshaking procedure and a bulk data encryption procedure. A client initiates a

connection with a server by sending a Client Hello message that includes the session ID, a random number, cipher suites, and other required information. After receiving the Client Hello, the server sends a Server Hello including its certificate and other information as a reply. With the certification of the server, the client finishes the authentication of the server. Depending on the server side configuration, the next procedure for the client authentication is optional. If it is requested, the client needs to send the certificate to the server for verification. After the authenticating procedures, the client generates session keys for the encryption and decryption of data. The session is identified by the session ID that is shared between the clients and server.

The main aim of this research is to reduce the pay load of the Web server. The server load balancing methods such as Random walk algorithm, Round Robin [RR] model, SSL-Session model and SSL-LB are reducing the server pay load. Random walk algorithm does not consider the load in the system and selects a random node and execute the processes with reduced throughput. The Round Robin scheduling does not acknowledge priorities, and does not allow out-of-order processing. SSL-Session may expire in a short duration of time.

III. SYSTEM MODEL

There are different servers introduced in geologically found spots and SSL-LB disseminates the clients demand to the suitable server and diminishes the compensation heap of the server. The SSL-LB has been displayed thusly as it were.

The incorporated Web and application server acts as takes after: The solicitations from client procedures are circulated by the distributor to the servers. When a server gets a demand, it opens a SSL association with the client. On the off chance that the server has the customer's past session data, it avoids the RSA lopsided key operation. As per the demand sort (static or element), the demand is prepared in an unexpected way. In the event that the demand is for static substance, it is adjusted by the Web server module. Else, it is sent to the application server module. The application server conjures the correspondence module to send the demand to another hub.

The below Figure 1 represents the SSL-LB system model. End-user requests are sent to a SSL-LB load-balancing system that determines which server is most capable of processing the request. It then forwards the request to that server. Server load balancing can also distribute workloads to firewalls and redirect requests to proxy servers and caching servers.
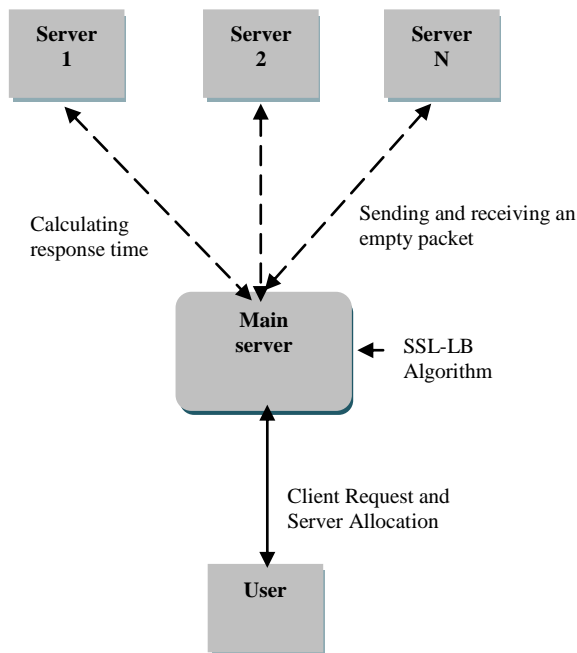

Figure 1: SSL-LB System Model

The Figure 2 represents outline of the load balancing system which is navigating the user requests to the sub-server when the main server is busy. Client's request sent to the load balancing system and load balancer calculate the response time for sub-servers. Based on the response time, the particular lightly loaded sub-server will be allocated to the client. This system has been modeled using various steps such as entry module, server representation, sharing resources and connecting clients with the sub-servers.
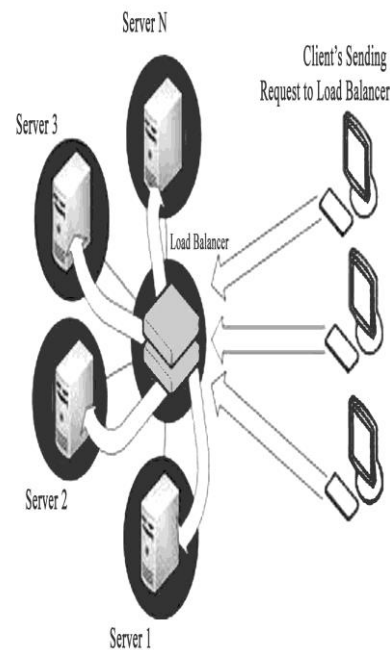

Figure 2: Outline of Load Balancing System

IV. SSL-LB IMPLEMENTATION

The source control program is operating similar to visual source safe. It performs sensibly easy by getting to the customer to set out refreshed data all the while. Generally it is completed utilizing Visual Studio.net or Front Page for assessment. At the point when every one of the procedures keep running on the group based server, the key capacity is load adjusting. Assume any one server in the cluster may harm, then the NLB cluster will re-balance the currently access request to continue the further operation. The principle favorable position of this strategy is that there is no connection between the servers in the cluster; consequently the impacts in one server won't influence the other server in the group.

Input text document is given by administrator and the scrambled arrangement is put away in all servers. The load balancing of the system technique will assembled a ranch of server. This server cultivate contains undesirable group getting to a few server benefit on a solitary IP address. Output module is a decrypted file from the server which is out of gear state. Yield module produces reaction time graph. The Network Load Balancer [NLB] handle the server related program and each approaching sign is acknowledged by every one of the hubs, however just the server with light load, will reaction rapidly. NLB apportions sub-server with less response time to the client who makes the demand.

Another advantage of the network load balancing scheme can gives idea of how much percentage of load balancing can be performed and find how the load can be shared within the nodes in the cluster. With the help of this load percentage calculation, it is very easy to select and transmit the data through the lightly loaded server. The user constantly distributed to the all nodes of the Web server, so that it is possible to find the percentage of load among the servers. In this case, the load of the server will change periodically during the request sending and receiving to the server. But in SSL-Session backend forwarding scheme, the load balance will be constant factor, hence there is no possibility of variance.

For applications, such as servers, this has numerous clients and relatively short-lived client requests. The ability of Network Load Balancer is to distribute workload efficiently and provides fast response to cluster changes. The network load balancing cluster will send the request to the all the node in the network, and looking for the response. Suppose the response did not receive from any other nodes, it shows the node will not function properly. From this idea, it is automatically re-arrange the server in the node. Every server in the cluster based application provides high level of security. Except database, it can be able to operate along and it does not need any other server for support. The server can be configured separately and execute the server operation without any help from other server. But in case of static Web server, it is not possible to run HTML file separately and there must be a problem raised when support file missing.

A. START SSL-LB

SSL-LB session is a logical connection between a client or server application over a Transmission Control Protocol (TCP) socket by using the SSL protocol. The SSL application creates a TCP socket, starts the TCP connection, and then starts the SSL-LB session over the TCP connection. The SSL-LB session is mapped to the TCP socket, therefore if the socket fails, the SSL-LB session will fail. When the SSL application sends data, the data is encrypted by the SSL code and sent across the TCP socket to the remote node. The SSL code in the remote node reads the data from the TCP socket, decrypts the data, and passes the data to the SSL application.

B. SSL-LB HANDSHAKE

SSL-LB session always begins with an exchange of messages called an SSL-LB handshake. The SSL-LB handshake allows the server to authenticate itself to the client by using public-key techniques. It should allow the client and server to cooperate in creating symmetric keys that are used for encryption, decryption, and tamper detection during the SSL-LB session that follows. The SSL-LB handshake can also allow the client to authenticate itself to the server.

C. END SSL-LB

To end SSL-LB session gracefully, it is ended before the Transmission Control Protocol (TCP) socket is ended. The client and server applications issue the SSL-shutdown function, which causes the SSL-LB session shutdown alert to flow across the TCP socket to the remote node. After the client and server applications have issued the SSL-shutdown function, the SSL-LB session is ended and the applications can then close the TCP socket

D. SSL-LB CONNECTION

1) Connection: It is client and server logic link; it associated with provision of suitable types of server. In SSL-LB terms, it must be a peer-to-peer connection.

2) Session: An association between client and server that define a set of parameter such as algorithm used.

The SSL-LB Session created by Handshake protocol that allows a parameter to be shared among the connection between client and server, and session is used to avoid negotiation of new parameters of each connection. A single session is shared among multiple SSL connection between client and server. SSL-LB session and connection involves several parameters that are used for SSL-LB enable communication between client and server. During the negotiations of the handshake protocol, the encryption methods are established and a series of parameters the session state are subsequently used within the session. Mostly SSL protects the HTTP communication channel over the internet and therefore the SSL protocol is seen quite often as associated only with www pages. The SSL protocol can be used to protect the transmission for any TCP/IP service.

V. CONCLUSION

Server load balancing is a powerful technique for improving application availability and performance in service provider, Web content provider and enterprise networks but implementation can also increase network cost and complexity. There are various servers introduced in geologically found spots and SSL-LB conveys the customer's demand to the suitable server and diminishes the compensation heap of the server. Extraordinary Networks gives the key advantages of server load balancing while dispensing with the potential cost, multifaceted nature, and execution issues.

VI. REFERENCE

[1] P Rafiq, J Kann: Network Load Balancing and Its Performance Measures, published on May 19,2015. International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 1, Jan-Feb 2015. ISSN: 2347-8578.

[2] Dipesh Gupta, Hardeep Singh: Review on TLS or SSL session sharing based web cluster load balancing. In proceedings of International Journal of Research in Engineering and Technology.Volume: 03 Issue: 11, Nov-2014. ISSN: 2321-7308.

[3] Web server load balancing using SSL back-end forwarding method V. M. Suresh; D. Karthikeswaran; V. M. Sudha; D. Murali Chandraseker IEEE – International Conferences on Advances in Engineering, Science and Management (ICAESM-2012).

[4] Jin-Ha Kim, 2007. An SSL Back-End Forwarding Scheme in Cluster Based Web Servers. IEEE transactions on parallel and distributed systems, 18(7):946-957. ISSN: 1045-9219.

[5] JH Kim, GS Choi, CR Das: A Load Balancing Scheme for Cluster-based Secure Network Servers, proceedings in IEEE International Conference On Cluster Computing, on September 2005, pp. 1-10.

[6] Shilpi pandey,Shivika prasanna : Load Balancing Techniques: A Comprehensive study published on April 2015,International journal of Advance research in Computer science and management studies, Volume 3,Issue 4,ISSN:2321-7782

1st International Conference on Applied Soft Computing Techniques
22 & 23.04.2017
In association with
International Journal of Scientific Research in Science and Technology

Dynamic Optimization Scheduling Techniques for Huge Data Centres in Cloud Computing Using QPSO Techniques

R. Sundarajan*[1], R. Arveena*[2]
*Associate professor, IT Department, KalasalingamUniversity, Virudhunagar, Tamilnadu, India
[2]PG Scholar, IT Department,Kalasalingam University,Virudhunagar, Tamilnadu, India

ABSTRACT

Load-balanced scheduling for huge server in clouds, in which a lot of information should be exchanged much of the time among a great many interconnected servers, is a key and testing issue. Existing Openflow based scheduling schemes, be that as it may, statically set up routes only at the initialization stage of data transmissions, which suffers from dynamical flow distribution and changing network states in data centers and often results in poor system performance. A novel dynamical load-balanced scheduling(DLBS) approach for boosting the system throughput while adjusting workload progressively. Here how we calculate the performance of the time delay and then optimize the performance of virtual machine. So in this process we optimizing dynamic scheduling and the process is how efficiently allocate the cloudlet in virtual machine using Quantum behaved particle swarm optimization (QPSO) to provide better and more efficient scheduling routing which is beneficial for both user and service provider. We used cloudsim tool to analyse how it optimized compared then previous result so far.

Keywords : Dynamic Load balancing, QPSO, Cloud Computing, Optimization

I. INTRODUCTION

Load balancing is a method to distribute workload across one or more interfaces, hardware devices, etc. It is a migration of a load from source node to lightly loaded destination node.

Load balancing approach may be static or dynamic. In the static algorithm, earlier learning about the system is as of now known which incorporates handling power, memory, execution and information about client's prerequisites. This calculation needn't bother with the data in regards to current condition of the system. Those having genuine level of disadvantages if there should arise an occurrence of sudden disappointment of framework asset assignments and furthermore errand can't be moved amid its execution for load adjusting.

Dynamic load balancing, when we apply load balancing while processing is called dynamic load balancing. Here there are two methods when an iterative method, final destination node estimated by several iterative stages. When an direct method, final destination node is selected in one step. Balance section is in charge of deciding instatement handle where virtual machine will begin. Work balancing is done in like manner and it gets the most extreme usage of the assets as per the information estimate in advance. A portion of the outcome demonstrates that specialists receiving the appropriated critical thinking methods are productive and successful in adjusting data centers and combining heterogeneous burdens. When we are considering the load balancing concept we have to remember some of the facts that contribute in it such as Throughput, Execution time, Delay time, processing time like that. Those are all plays important role in cloud computing concept as well as performing in the datacenter. Either of these patterns might deal with huge amount of data, but a storage solution would be hard to handle efficiently while treating their data undifferentiated.

Global optimization:

Optimization, in a general sense, has the objective of getting the most ideal outcome given a scope of

decisions. These decisions can be spoken to with factors in a capacity, and the outcome represented by the capacity assessment. Consequently lead to a given capacity is to look for the parameters which prompt to the biggest, or littlest, conceivable result. Regardless of whether the biggest or littlest esteem is sought relies on upon the specific application, however for either case; the issue can essentially be turned to produce the other. This permits all issues to be dealt with as minimization issues, which will be the situation for the optimized result.

Selecting the best subset
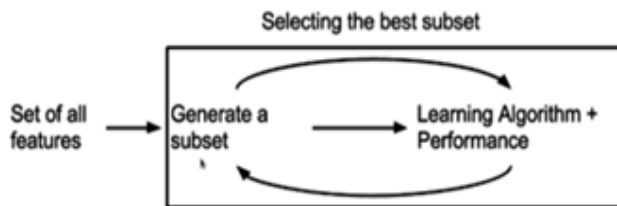
II. RELATED WORK

Before In earliest decades, noteworthy consideration has been dedicated to the task assignment and load adjusting in conveyed frameworks. Despite the fact that there have been some related overviews about this subject, each of which just made an extremely preparatory survey on the condition of specialty of one single sort of circulated framework. To associate the reviews in changing sorts of circulated frameworks and make an exhaustive scientific categorization on them, this study basically arranges and audits the delegate thinks about on Task allocation and load balancing as per the general qualities of fluctuating disseminated frameworks. To begin with, this study condenses the general qualities of characteristics of distributed and dynamic load balancing system. Based on these general qualities, this overview surveys the reviews on task assignment and work load balancing as for the accompanying perspectives: 1) ordinary control models; 2) ordinary resource optimization method 3) Strategies for accomplishing dependability; 4) common coordination mechanism among heterogeneous nodes; and 5) typical models considering network structures. For every perspective, we compress the current reviews and talk about the future research based upon the

previous result. Through the overview, the related reviews here can be surely known based of how they can fulfill the general qualities of appropriated frameworks.

In this previous research, they address the load-balanced scheduling problem through balancing transmission traffic dynamically and globally in cloud data centers. Aiming at two typical OpenFlow architectures: FPN and FTN, they proposed and implemented a set of efficient scheduling algorithms DLBS-FPN and DLBS-FTN respectively. Compared with previous scheduling schemes for load balancing and path selection, those algorithms have two main advantages. Firstly, our algorithms can adapt to dynamical network states and changing traffic requirements through updating load imbalance factor $\delta(t)$ and accordingly balancing the transmission load slot by slot during data transmissions. Next, our algorithms can globally balance transmission traffic in the whole network by means of evaluating link, path and network bandwidth utilization ratio proposed in this paper.

In synopsis, they can find that their DLBS shows a higher data transmission usage rate than LOBUS and RR in three transmission designs during the underlying stage, and conveys more transmission load before comparing basic time.

In this way, they can infer that their DLBS can effectively adjust the worldwide load so that it altogether enhances throughput, transmission delay and data transfer capacity use rate particularly under non-uniform system transmission designs.

Both of DLBS approach are mainly used other scheduling algorithms to improve their efficiency to shortly the response from the user herewith we consider the throughput.

III. ALGORITHM DESCRIPTION

After Particle Swarm Optimization (PSO) is moderately a more up to date expansion to a class of population based scan method for tackling numerical streamlining issues. The particles or individuals from the swarm fly through a multidimensional scan space searching for a potential arrangement.

Apply DTA Algorithm

In this module the data can be analysis for every process related to the cloud and then the time can be calculated related to the data centre analysis and then the related broker node can be created. Then the allocation of the data to the virtual machine can be processed related to the time analysis.
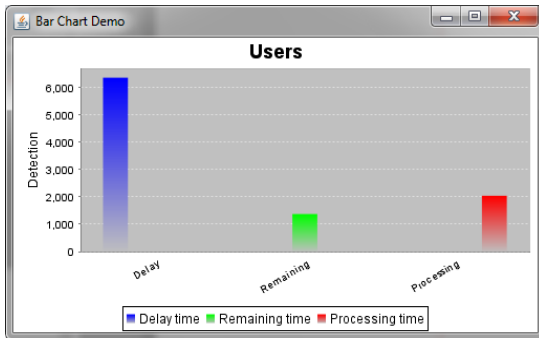
Figure 1.1 Time chart occur in the process

After the Evaluation of Delay Time Algorithm we found out the time delay that occurs in our job to allocate the schedule in server. Based upon the analysis we have to give the optimisation to improve the performance. Fig 1.1 shows the time occur in each cloudlet to process their job.

One of the current advancements in pso is the utilization of Quantum laws of mechanics to watch the conduct of PSO. A few variations of QPSO incorporate change based PSO, where transformation is connected to the Mbest (mean best) and Gbest (worldwide best) places of the molecule, likewise in one of the variations of QPSO an irritation steady is included. Besides to the best of our insight no one has utilized the idea of recombination administrator in QPSO.

Quantum-carried on molecule swarm advancement (QPSO) calculation is a worldwide joining ensured calculation, which beats unique PSO in pursuit capacity yet has less parameter to control. In our work, we recommended a quantum based particle swarm optimisation with highest weighted mean good position according to minimum fitness values of positions. It is clearly understand that the QPSO has good performance speed, resulting in superior balance between the local and global seek algorithm, and thus producing better performance outcome. The implemented QPSO is tested

on several capabilities and then compared with PSO and then with Delay time Algorithm Particle Swarm Optimization (PSO) is moderately a more up to date expansion to a class of population based scan method for tackling numerical streamlining issues. The particles or individuals from the swarm fly through a multidimensional scan space searching for a potential arrangement.

IV. PSEUDOCODE

The computational steps of QPSO algorithm are given by:

Step 1: Instate the swarm with consistently appropriated irregular numbers.

Step 2: Find out mbest using formula.

Step 3: Update particle position using fitness equation.

Step 4: Assess the fitness estimation of every molecule and then find minimum fitness.

Step 5:In the event that the present fitness esteem is superior to the best fitness esteem (Pbest) in history

At that point Update Pbest by the present fitness esteem.

Step 6: Modified gbest (global best)

Step 7: *Determine a new value using that equation.*

Step 8: *In the event that the new particle is superior to the speculative particle in the group which having minimum fitness value then replace the particle having minimum fitness function.*

Step 9: Go to further until maximum iteration be reached.

In the above calculation, in the quantum model of a PSO, the condition of a molecule is delineated by wave function Y(x, t), rather than position and speed. The alternate conduct of the molecule is generally different from that of the molecule in customary PSO frameworks. In this unique situation, the likelihood of the molecule's showing up in position xi from likelihood thickness function.

Mean best (mbest) of the corpuscle is characterized by mean of the superior places of all particles, are consistently circulated irregular values in the interim [0, 1]. The parameters are called compression development coefficient. The flow chart of recommended QPSO used in our domain are shown in fig 1.3


```
Initialize the Swarm
Do
Calculate mbest by equation (5)
Update particles position using equation (3)
Update Pbest
Update Pgbest.
While maximum iteration is reached.
```

Figure 1.2 Pseudo code of QPSO

V. Flow diagram

It shows the process of QPSO that should be done in dynamic load balancing and the step by step process that should be applied in the implementation of this paper. Fig 1.3 shows the flow diagram of Quantum based particle swarm optimization.
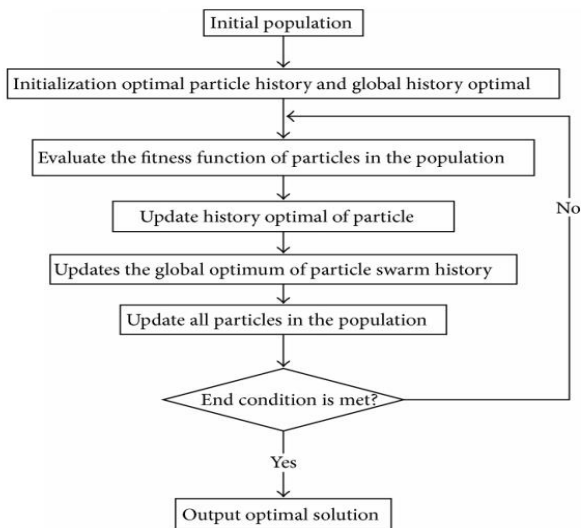

Figure 1.3 Flow chart of proposed QPSO

VI. Experimental results

The implementation of Delay time Algorithm is to calculate the processing time, Delay time and remaining time in the scheduling process.
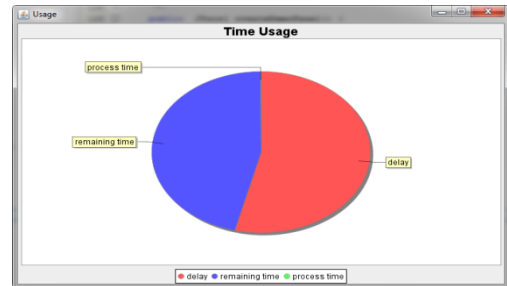

Figure 1.4 Time executed in the process

As we immerge Q PSO algorithm in Dynamic Load balancing for efficient retrieve of server response at the same time give optimised result for the load balancing scheduling approach in cloud computing concept.

The performance of QPSO in load balancing technique is compared with other optimisation algorithm and got best result when compared. It improves the performance and decrease the execution period of the task. Fig 1.5 shows the synthetic traffic that occurred in the virtual machine with multiple jobs taken.
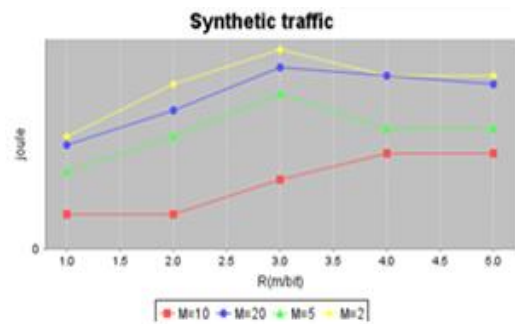

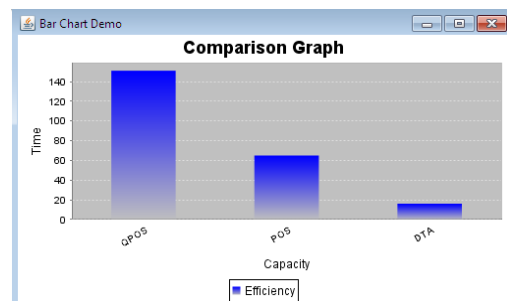**Figure 1.5** Traffic occurred in the process

Figure 1.6 Comparison graph of utilised algorithm

VII. CONCLUSION

Task assignment and work load balancing have been seriously inquired about in past decades; an extensive number of related reviews and outcome have been exhibited as regards of this subject. First of all we consider load balancing and then to optimised dynamic load balancing DLBS approach for maximizing the network throughput while balancing the workload dynamically. The process is under how efficiently allocate the cloudlet in virtual machine using Quantum based particle swarm optimisation algorithm. It showed their efficiency while performing their optimisation.

Compared with other traditional algorithms, the experiment results on standard testing functions showed that the recommended algorithm has the superior optimization ability, with enhancement in both convergence speed and time execution those demonstrate the effectiveness of the QPSO that applied in load balancing concept.

VIII. REFERENCES

[1] Feilong Tang *Member, IEEE*, Laurence T. Yang *Senior Member, IEEE*, Can Tang, Jie Li *Senior Member, IEEE*, and Minyi Guo *Senior Member, IEEE* A Dynamical and Load-Balanced FlowScheduling Approach for Big Data Centers inCloud.

[2] Luyu Wang, Shiyou Yang, Jin Huang, Fumio Kojima, Futoshi Kobayashi, Hiroyuki Nakamoto. (2016) An adaptive quantum-behaved particle swarm optimizer for global optimization of inverse problem. *International Journal of Applied Electromagnetics and Mechanics* **52**:1-2793-799. Online publication date: 29-Dec-2016

[3] Obaid Ur Rehman, Shiyou Yang, Shafi Ullah Khan. (2017) A modified quantum-based particle swarm optimization for engineering inverse problem. *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering* **36**:1168-187.

[4] Z.Z.Cao, M.Kodialam and T.V.Lakshman. Joint Static and Dynamic Traffic Scheduling in Data Center Networks. in Proceedings of *IEEE INFOCOM 2014, pp.2445-2553*

[5] Maolong Xi, Xiaojun Wu, Xinyi Sheng, Jun Sun, Wenbo Xu. (2016) Improved quantum-behaved particle swarm optimization with local search strategy. *Journal of Algorithms & Computational Technology*174830181665402.

[6] Chia-Yu Wang, Pei-Rong Li, Chia-Lin Tsai, Kai-Ten Feng. (2016) Load-balanced user association and resource allocation under limited capacity backhaul for small cell networks. *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications(PIMRC)*1-5.

[7] Tianyu Liu, Licheng Jiao, Wenping Ma, Ronghua Shang. (2017) Quantum-behaved particle swarm optimization with collaborative attractors for nonlinear numerical problems.

[8] Tianyu Liu, Licheng Jiao, Wenping Ma, Ronghua Shang. (2017) Quantum-behaved particle swarm optimization with collaborative attractors for nonlinear numerical problems.

[9] Yang Wang, Yangyang Li, Licheng Jiao. (2016) Quantum-inspired multi-objective optimization evolutionary algorithm based on decomposition. Soft Computing 20:83257-3272.

[10] Jui-Yu Wu An improved quantum-behaved particle swarm optimization method for solving constrained global optimization problems, 15th International Symposium on Communications and Information Technologies(ISCIT)Year:2015

[11] Quantum-Behaved Particle Swarm Optimization with Cooperative Coevolution for Large Scale Optimization Na Tian2015 *14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*Year: 2015

[12] Tianyu Liu, Licheng Jiao, Wenping Ma, Ronghua Shang. (2017) Quantum-behaved particle swarm

optimization with collaborative attractors for nonlinear numerical problems.

[13] Gulshan Soni; Mala Kalra A novel approach for load balancing in cloud datacenter 2014 *IEEE International Advance Computing Conference (IACC)Year:2014.*

[14] Hybrid Memetic and Particle Swarm Optimization for Multi Objective Scientific Workflows in CloudK. Padmaveni; D. John Aravindhar2016 *IEEE International Conference on Cloud Computing in Emerging Markets(CCEM)Year:2016*