Knowledge Exploration in Image Text Data using Data Hiding Scheme

M. Thangamani, *Member, IAENG,* Jafar Ali Ibrahim. S, *Member, IAENG*

Abstract: **Reversible data hiding is a technique used to hide secret message into a cover image by modifying its pixel values slightly. The cover image and embedded message are completely recovered from the marked content. It helps in information hiding with the lossless compressibility of natural images. Histogram modification, difference expansion, Lossless compression, prediction-error expansion and integer transform techniques are used to carry out this research. In this method which is based on histogram is divided into two stages: Histogram generation and Histogram modification. Histogram construction is performed with the pixel pair arrangements and their dissimilar values. Histogram modification is carried out to embed data into the cover image. The process of unhiding recovers the secret message and also the cover image.**

Index Terms**—Reversible Data Hiding, image mining, Stego image, Multiple histogram and RSA algorithm**

I. INTRODUCTION

Prediction-error expansion (PEE) technique is applied for Reversible Data Hiding (RDH) process. One-or two-dimensional Prediction-error Histogram (PEH) are used in the PEE techniques. The two-dimensional PEH-based methods accomplish better than one dimensional PEH. PEH alteration is fixed and independent of image content. Multiple histograms based PEE method is accepted to improve reversible data hiding process. Multiple Histograms Modification (MHM) method uses a sequence of histograms for the hiding process. A complexity measurement is computed for each pixel with reference to its context. Prediction-Error Histogram (PEH) is generated using the pixels with the complexity value. A sequence of histograms can be prompted by varying the difficulty to cover the full image. Two expansion bins are selected in collectively generated histogram. The expansion bins are carefully chosen with reference to the image content.

The Multiple Histogram Modification (MHM) model is enhanced to increase the embedding capacity. Predictor selection process is improved in the hiding process. Complexity measurement selection process is customized in the MHM process. Histogram modification is implemented with different frequency levels. RSA algorithm is applied on the secret data to increase the security levels.

Dr. M. Thangamani is Assistant Professor, Kongu Engineering College, manithangamani2.gmail.com

Mr. Jafar Ali Ibrahim S is Doctorial Scholar, Anna University, jafartheni@gmail.com

Reversible data hiding (RDH) aims to embed secret message into a cover image by slightly modifying its pixels and more importantly, the original image as well as the embedded message should be completely restored from the marked image. RDH has received much consideration from the information hiding community and this technique has also been applied in some applications such as image authentication, medical image processing, multimedia archive management, image trans-coding and data coloring in the cloud, etc. RDH is a fragile method and the marked image cannot go through any degradation. RDH method is usually evaluated by its capacity-distortion performance, i.e., for a given embedding capacity (EC), one expects to minimize the embedding distortion measured by PSNR of the marked image versus the original one.

II. RELATED WORK

Early RDH methods are mainly based on lossless compression. The idea behind these approaches is to compress without loss a feature set of cover image and use the saved space for reversible embedding. Fridrich et al.[1] proposed to compress a proper bit-plane with the minimum redundancy. Celik et al. [2] projected a generalized least significant bit (LSB) compression method to increase the compression efficiency by means of unaltered bit-planes as side information. The lossless compression-based methods cannot produce acceptable performance, since the relationship within a bit-plane is too weak to afford a high EC. As EC increases, one needs to compress more bit-planes, thus the distortion increases dramatically. In the system [3], the RP-based methods are extended by applying a variety of embedding schemes. Here, the concept of dual binary tree has been applied for the embedment of data which helps in increasing the payload to a greater extent. The predictor has been used to obtain the predictor errors and an energy estimator has been used to calculate the energy of each prediction error. Thus, the embedding efficiency has been greatly increased. In the histogram-shifting method, initially, each local image region is projected to a one-dimensional space to obtain a scalar sequence which is followed by the generation of one-dimensional histogram. Secondly, the data are implanted into the cover image by transforming the histogram. A local image region has been considered and then it is projected to a two dimensional space to get a sequence which consists of difference-pairs. Then, a two-dimensional difference-

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

histogram is generated by counting the difference-pairs [4]. Here more number of pixels has been used for carrying data with the significant reduction in the number of shifted pixels and significant improvement in the embedding mechanism.

Video Steganography falls under two categories: The one being the embedment of the secret message in the pixels of the video frames directly [5]. The other one hides the secret message into the compressed domain of the videos by modifying the variable length codes, prediction modes and the motion vectors. For video Steganography in spatial domain, in order to guarantee the successful recovery of hidden message after potential distortion due to lossy compression or channel noise, it usually encodes the message with error correcting code and embeds them into multiple locations, which resists the video compression to some degree but limits the embedding capacity. The other category is specially designed for compressed video. It directly modifies the coefficients in the compressed domain, thus it integrates Steganography directly into the compression scheme but is fragile to any distortion and the embedding capacity is also limited because there are fewer coefficients which can be modified to carry the secret message.

The stego image with a low embedding ratio is usually detected more difficultly, the steganalysis of the stego image with a low embedding ratio has been one of the important and difficult issues of steganalysis [6]. Although the structural steganalysis can estimate the low embedding ratio with less error than others, the existing structural steganalysis methods for MLSB steganography only utilize the correlation between two adjacent samples. For LSB replacement, the structural steganalysis methods utilizing the correlation among more than two adjacent pixels can usually estimate the low embedding ratio with higher precision. Intuitively, one should also be able to improve the estimation accuracy for the low embedding ratio of MLSB steganography by utilizing the correlation among more adjacent pixels. Reversible data hiding techniques can be done in spatial domain as well as in compressed domain. In spatial domain embedding, the data are embedded into the pixels through modifying pixel values in the cover image [7]. This technique can achieve larger payload since the images in spatial domain provides rich redundancies, which are suitable for data embedding. In the compressed domain embedding, the embedding mechanism is carried out by modifying the compressed codes.

In the proposed method, the LSBs of some pixels are embedded into other pixels with a conventional RDH method which paves the way for the system to empty out the room followed by the encryption of the image [8]. Thus the data implanting is accomplished by using the position of the LSBs. In this method, excellent performance is achieved in two steps: Reversibility is achieved to a maximum extent since the extraction mechanism and recovery of images are error free. Real reversibility is recognized, that is, data extraction and image recovery are free of any error. And the Peak Signal to Noise Ratio is highly improved. certain dimension. It has two major advantages. On one hand, the maximum modification to pixel values can be controlled and thus the embedding distortion can be well limited [9]. On the other hand, the location map used to record underflow/overflow locations is usually small in size especially for low ER case. As pointed out by Sachnev et al., RDH algorithm with smaller, or in some cases, no location maps, are very desirable. Therefore, HS is a good choice among existing approaches of RDH. Reversible image watermarking, which enables recovering both the original image and the watermark from the watermarked content, has aroused considerable interest recently. Many valuable reversible watermarking algorithms have been presented in the literature. Tian introduced a difference expansion (DE) based method, in which the difference of two adjacent pixels is expanded to carry one data bit [10]. This method usually provides high capacity while keeping the distortion low.

III. PROPOSED ARCHITECTURE

Reversible Data Hiding (RDH) techniques are used to support data hiding with message and cover image retrieval mechanism. Multiple Histogram Modification (MHM) scheme is employed for the data hiding process. The system is enhanced to improve the embedding capacity with optimized predictor selection approach. RSA algorithm is adapted to ensure the security level of secret data values. The Multiple Histogram Modification (MHM) model is enhanced to increase the embedding capacity. Predictor selection process is improved in the hiding process. Complexity measurement selection process is customized in the MHM process. Histogram modification is carried out with different frequency levels. RSA algorithm is applied on the secret data to increase the security levels.

The data hiding scheme is enhanced to distribute the secret text to the required users. The hiding process is carried out on the cover image values. The secret text is secured using the Cryptography techniques. The encryption operations are carried out under the sender environment. The decryption operations are carried out in under the receiver environment. The public key is collected from the receiver and it is used to encrypt the secret text values. The encrypted text is passed into the data hiding process. The steago image is transferred to the receiver node. The receiver node collects the steago image and performs the unhide operations. The decryption process is called after the completion of the unhide operations. The system uses the Multiple Histogram Modification (MHM) mechanism. The RSA algorithm is used for the data security process.

The data selection process is designed to handle the secret text and cover image selection tasks. The secret text path and its contents are assigned for the hiding process. The cover image selection is carried out the support dialog boxes. Image properties are extracted by the system. Image name, image type, size, height and width properties are identified by the system. The key management process is designed to generate the public and private key pairs. All the secret text, cover image and key management parameters are passed to the data transmission process. The received images are stored under the downloads folder in the receiver system.

The RSA algorithm is applied to secure the secret data value. The system is divided into four major modules. They are Sender, Data security, and Receiver and Data extraction.

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The sender module is designed to send the image data values. Data security is designed to perform hiding and encryption process. Data receiver module collects data from the sender. Data extraction module is designed to perform unhide operations.

Sender: The data sender collects secret data and cover data from the user. Prediction-Error Histogram (PEH) technique is used in the system. Multiple Histogram Modification (MHM) is applied with Prediction Error Histogram (PEH) with different complexity levels. Histogram construction is carried out using the cover data image.

Data Security: Data security process is used to hide secret data values. Secret data is converted into bits. RSA algorithm is used to encrypt the data values. The sender collects the public key from the receiver node for the encryption process. Encrypted data values are hided in the cover image.

Receiver: Data receiver collects data from the sender node. Received data values are updated into the local memory. The received data value is passed to unhide and decrypt process. The receiver node maintains the secret key for decryption process.

Data Retrieval: Secret data is separated from the received data values. Cover data is also separated from the received data values. Decryption process is carried out to fetch the secret data. Cover data quality is analyzed with image quality measures.

IV. RESULTS AND DISCUSSION

The Reversible Data Hiding (RDH) scheme is improved with Cryptography techniques and optimal difference pair mapping process. The secret text values are embedded in the cover image under the encoding process. The secret text is encrypted before the encoding process. The RSA algorithm is used for the encryption and decryption process. The decryption process is carried out after the decoding process. The system is designed as sender and receiver application. The encryption and encoding operations are performed in the sender application. The decoding and decryption operations are carried out under the receiver process. The sender transforms the embedded image to the receiver. The receiver decodes the image and extracts the secret text.

The secured data transmission system is tested with sender and receiver application. Reversible Data Hiding with Multiple Histogram Modification (RDH-MHM) model and Reversible Data Hiding with Optimized Multiple Histogram Modification (RDH-OMHM) models are used in the system. The RDH-MHM model performs the data hiding on histogram points. The RDH-OMHM model performs the data hiding process on optimized DPM positions. The system selects the threshold for the histogram dynamically. The RDH-OMHM scheme also uses the encrypted message for the hiding process. The system is tested with different set of data transmission cycles. The system performance is measured with three parameters. They are Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE) and Embedding Capacity (EC) metrics. The PSNR and MSE metrics are used to evaluate the received image quality. The hiding capacity is analyzed using the embedding capacity measure.

Peak Signal-to-Noise Ratio (PSNR)

The image quality is analyzed using Peak Signal-to-Noise Ratio (PSNR) method. Peak Signal-to-Noise Ratio is the ratio between the reference signal and the distortion signal in an image, given in decibels. The higher the PSNR, the closer the distorted image is to the original. In general, a higher PSNR value should correlate to a higher quality image, but tests have shown that this is nott always the case. PSNR is a popular quality metric because it is easy and fast to calculate while still giving okay results. For images A = {a1 ..aM}, B = {b1 .. bM} and MAX equal to the maximum possible pixel value (2^8 - 1 = 255 for 8-bit images):

$$PSNR(A,B) = 10log_{10}\left(\frac{MAX^2}{MSE(A,B)}\right)$$

Mean Squared Error (MSE)

The image quality analysis for the data hiding and transmission process is performed using the Mean Square Error (MSE) metric. Mean Squared Error is the average squared difference between a reference image and a distorted image. It is computed pixel-by-pixel by adding up the squared differences of all the pixels and dividing by the total pixel count. For images A = {a1 ..aM} and B = {b1 .. bM}, where M is the number of pixels:

$$MSE(A,B) = \frac{1}{m}\sum_{i=1}^{M}(a_i - b_i)^2$$

Embedding Capacity (EC)

The performance of data hiding schemes is verified with Embedding Capacity (EC) parameter. Embedding Capacity is used to represent the amount of data that can be hided in the cover image. The embedding capacity is estimated with the possible pixel count that can be obtained for the hiding process. In the Reversible Data Hiding (RDH) scheme the pixels that are selected for the hiding process is denoted as embedding capacity. The pixel selection is performed with the histogram analysis mechanism. The Reversible Data Hiding with Multiple Histogram Modification (RDH-MHM) scheme uses the fixed threshold for the pixel selection process. The Reversible Data Hiding with Optimal Multiple Histogram Modification (RDH-OMHM) system uses the dynamic threshold value for the hiding method.

V.RESULT INTERPRETATIONS

The Reversible Data Hiding with Multiple Histogram Modification (RDH-MHM) pattern and Reversible Data Hiding with Optimized Multiple Histogram Modification (RDH-OMHM) pattern are verified using Peak Signal to Noise Ratio (PSNR) metric for image quality analysis in receiver end. PSNR analysis is shown in Fig. 1 and table 1. The analysis displays that the PSNR in RDH-ODPM is 15% increased than the RDH-DPM method.

Table.1 Image Quality Analysis using PSNR between RDH-MHM and RDH-OMHM Techniques

S. No	Images	RDH-MHM	RDH-OMHM
1	10	67.36	82.31
2	20	69.92	83.28
3	30	72.17	85.22
4	40	74.86	87.18
5	50	77.21	89.68

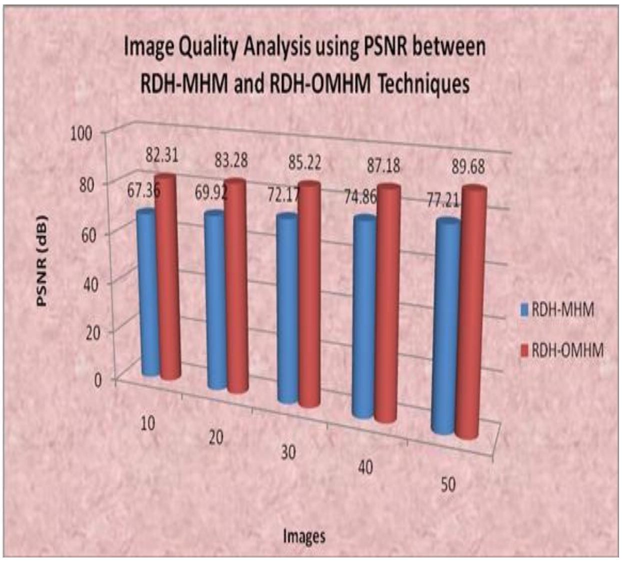

Fig.1 Image Quality Analysis using PSNR between RDH-MHM and RDH-OMHM Techniques

The squaring of the differences dampens minor changes between the 2 pixels but penalizes large ones. The Reversible Data Hiding with Multiple Histogram Modification (RDH-MHM) system and Reversible Data Hiding with Optimized Multiple Histogram Modification (RDH-OMHM) system are verified using Mean Square Error (MSE) metric for image quality analysis in receiver end. MSE analysis is shown in Fig.2 and Table 2. The analysis displays that the MSE in RDH-OMHM is 25% reduced than the RDH-MHM.

Table.2 Image Quality Analysis using MSE between RDH-MHM and RDH-OMHM Techniques

S. No	Images	RDH-MHM	RDH-OMHM
1	10	3.24	2.37
2	20	3.31	2.48
3	30	3.46	2.62
4	40	3.59	2.75
5	50	3.51	2.86

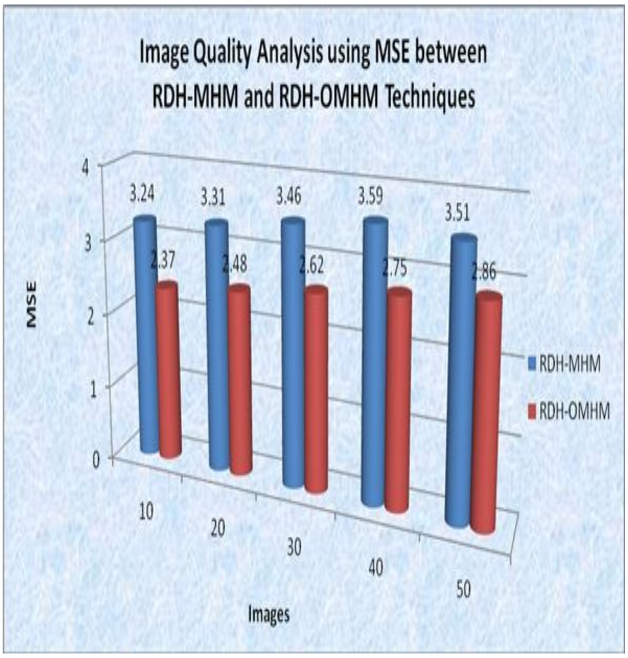

Fig. 2 Image Quality Analysis using MSE between RDH-MHM and RDH-OMHM Techniques

The squaring of the differences dampens minor changes between the 2 pixels but penalizes large ones. The Reversible Data Hiding with Multiple Histogram Modification (RDH-MHM) system and Reversible Data Hiding with Optimized Multiple Histogram Modification (RDH-OMHM) system are verified using Mean Square Error (MSE) metric for image quality analysis in receiver end. MSE analysis is shown in Fig.2 and Table 2. The analysis displays that the MSE in RDH-OMHM is 25% reduced than the RDH-MHM.

The Reversible Data Hiding with Multiple Histogram Modification (RDH-MHM) system and Reversible Data Hiding with Optimized Multiple Histrogram Modification (RDH-OMHM) system are verified using Embedding Capacity (EC) parameter for hiding performance analysis. Embedding Capacity (EC) parameter analysis is shown in Figure 4 and Table 3. The analysis displays that the embedding capacity in RDH-OMHM is 30% improved than the RDH-DMHM.

Table No.3 Embedding Capacity (EC) Analysis between RDH-MHM and RDH-OMHM Techniques

S. No	Images	RDH-MHM	RDH-OMHM
1	10	1623	2356
2	20	1697	2476
3	30	1783	2579
4	40	1830	2699
5	50	1887	2843

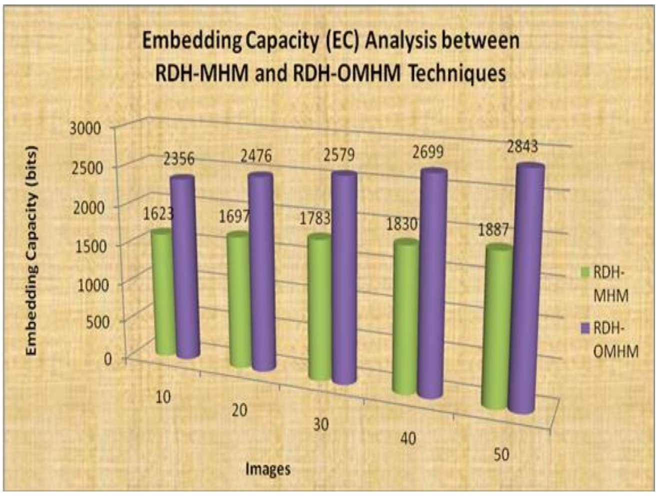

Fig.4 Embedding Capacity (EC) Analysis between RDH-MHM and RDH-OMHM Techniques

VII. CONCLUSION & FUTURE ENHANCEMENT

Reversible Data Hiding (RDH) techniques are used to support data hiding with message and cover image retrieval mechanism. Multiple Histogram Modification (MHM) scheme is employed for the data hiding process. The system is enhanced to improve the embedding capacity with optimized predictor selection approach. RSA algorithm is adapted to ensure the security level of secret data values. The MHM scheme is enhanced to improve the embedding capacity. Embedding performance is improved by the system. The system supports efficient coverage image retrieval process. The system reduces the process time in hiding and un-hiding process.

The reversible data hiding scheme based data distribution system can be enhanced with data integrity techniques to verify the data transfer process. The system can be enhanced to support intrusion detection process. The data hiding process can be applied on additional cover data mediums such as audio and video files.

REFERENCES

[1] Faxin Yu, Zheming Lu, Ha-o Luo, Pinghui wang, "Three dimensional Model analysis and processing", pp.374-378, 2011.

[2] Celik Mehmet Utku, Gaurav Sharma, Ahmet Murat Tekalp, Eli Saber, "Lossless Generalized-LSB Data Embedding", *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp.253-266, Feb. 2005

[3] Wien Hong, "Adaptive reversible data hiding method based on error energy control and histogram shifting", *Elsevier, Optics Communications*, vol. 285, no.2, pp.101-108, 2012.

[4] Xiaolong Li, Weiming Zhang, Xinlu Gui, and Bin Yang, "A Novel Reversible Data Hiding Scheme Based on Two-Dimensional Difference-Histogram Modification", *IEEE Transactions on Information forensics and Security*, vol. 8, no.7, pp.1091-1100, July 2013.

[5] Keren Wang, Hong Zhao and Hongxia Wang, "Video Steganalysis Against Motion Vector-Based Steganography by Adding or Subtracting One Motion Vector Value", *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 741-751, May 2014.

[6] Chunfang Yang, Fenlin Liu, Xiangyang Luo, and Ying Zeng, "Pixel Group Trace Model-Based Quantitative Steganalysis for Multiple Least-Significant Bits Steganography", *IEEE Transactions on Information Forensics and Security*, vol.8, no.1, pp.216-228, Jan. 2013.

[7] Wien Hong, Tung-shou chen, "A local variance-controlled reversible data hiding method using prediction and Histogram-shiefting", *Elsevier, Journal of systems and Software*, vol.83, no.12, pp.2653-2663.

[8] Kede Ma, Weiming Zhang, Xianfeng Zhao, Member, IEEE, Nenghai Yu, and Fenghua Li, "Reversible Data Hiding in Encrypted Images by Reserving Room Before Encryption", *IEEE Transactions on Information Forensics and Security*, vol.8, no.3, pp.553-562, March 2013.

[9] Xiaolong Li, Bin Li, Bin Yang, and Tieyong Zeng, "General Framework to Histogram-Shifting-Based Reversible Data Hiding", *IEEE Transactions on Image Processing*, vol.22, no. 6, pp.2181- 2191, June 2013.

[10] Xiang Wang, Xiaolong Li, Bin Yang, and Zongming Guo, Member, IEEE, "Efficient Generalized Integer Transform for Reversible Watermarking", *IEEE Signal Processing Letters*, vol.17, no.6, pp.567-570, June 2010.

AUTHORS BIOGRAPHY

Thangamani completed her B.E., from Government College of Technology, Coimbatore, India. She completed her M.E in Computer Science and Engineering from Anna University and PhD in Information and Communication Engineering from the renowned Anna University, Chennai, India in the year 2013.

Dr. M. Thangamani possesses nearly 23 years of experience in research, teaching, consulting and practical application development to solve real-world business problems using analytics. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing, ontology development, web services and open source software. She has published nearly 80 articles in refereed and indexed journals, books and book chapters and presented over 67 papers in national and international conferences in above field. She has delivered more than 60 Guest Lectures in reputed engineering colleges and reputed industries on various topics. She has got best paper awards from various education related social activities in India and Abroad. She has organized many self-supporting and government sponsored national conference and Workshop in the field of data mining, big data and cloud computing. She has received the many National and International Awards. She continues to actively serve the academic and research communities and presently guiding nine Ph.D Scholars under Anna University. She is on the editorial board and reviewing committee of leading research SCI journals. She has Editor-In-Chief in International Scientific Global Journal in Engineering Science and Applied Research (ISGJESAR) and on the program committee of top international data mining and soft computing conferences in various countries. She is also

Board member in Taylor & Francis Group and seasonal reviewer in IEEE Transaction on Fuzzy System, international journal of advances in Fuzzy System and Applied mathematics and information journals. She has organizing chair and keynote speaker in international conferences in India and countries like California, Dubai, Malaysia, Singapore, Thailand and China. She has Associate Editor in Canadian Arena of Applied Scientific Research, Canada. She has Life Membership in ISTE, Member in CSI, IETE, IAENG, IRES, Athens Institute for Education, Research and Life member in Analytical Society of India and annual member in IEEE. She is currently working as Assistant Professor at Kongu Engineering College at Perundurai, Erode District.

Jafar Ali Ibrahim. S has completed his Bachelor of Technology in Information Technology from Syed Ammal Engineering College, Ramanathapuram, affiliated to Anna University, Chennai, Tamilnadu, India and Master of Technology degree in Computer Systems and Networks from Dr MGR Educational and Research Institute, Chennai, Tamil Nadu, India. He has presented his papers in three International journals. His research interests include Cloud Computing, Big Data and Data mining.

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)