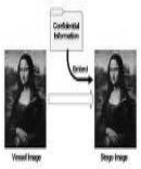
Volume : IV Issue No.1

CONTENT

WWW.ksrcejca.webs.com

Software Incubator for Budding Professionals



Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

Edge Adaptive Image Steganography Based On LSB Matching Revisited

[1]Mrs. Sivaranjani [2]Ms. Semi Sara mani

Abstract— **The least-significant-bit (LSB)-based approach is a popular type of steganographic algorithms in the spatial domain. However, we find that in most existing approaches, the choice of embedding positions within a cover image mainly depends on a pseudorandom number generator without considering the relationship between the image content itself and the size of the secret message. Thus the smooth/flat regions in the cover images will inevitably be contaminated after data hiding even at a low embedding rate, and this will lead to poor visual quality and low security based on our analysis and extensive experiments, especially for those images with many smooth regions. In this paper, we expand the LSB matching revisited image steganography and propose an edge adaptive scheme which can select the embedding regions according to the size of secret message and the difference between two consecutive pixels in the cover image. For lower embedding rates, only sharper edge regions are used while keeping the other smoother regions as they are. When the embedding rate increases, more edge regions can be released adaptively for data hiding by adjusting just a few parameters. Keywords-**

Index Terms **— Content-based steganography, least-significant-bit (LSB)-based steganography, pixel-value differencing (PVD), security, steganalysis.**

I. INTRODUCTION

STEGANOGRAPHY is a technique for information hiding. It aims to embed secret data into a digital cover media, such as digital audio, image, video, etc., without being suspicious. On the other side, steganalysis aims to expose the presence of hidden secret messages in those stego media. If there exists a steganalytic algorithm which can guess whether a given media is a cover or not with a higher probability than random guessing, the steganographic system is considered broken. In this paper, we consider digital images as covers and investigate an adaptive and secure data hiding scheme in the spatial least-significant-bit (LSB) domain. LSB replacement is a well-known steganographic method. In this embedding scheme, only the LSB plane of the cover image is overwritten with the secret bit stream according to a pseudorandom number generator (PRNG). As a result, some structural asymmetry (never decreasing even pixels and increasing odd pixels when hiding the data) is introduced, and thus it is very easy to detect the existence of hidden message even at a low embedding rate using some reported steganalytic algorithms.

Manuscript received May 14, 2011.

Mrs. Sivaranjani[1], Department of Computer Science and Engineering, Avinashlingam Deemed University for Women, Coimbatore.
Ms. Semi Sara mani[2], PG scholar, Department of information technology, Easa College of engineering & technology, Coimbatore.

LSB matching (LSBM) employs a minor modification to LSB replacement. If the secret bit does not match the LSB of the
cover image, then or is randomly added to the corresponding pixel value. Statistically, the probability of increasing or decreasing for each modified pixel value is the same and so the obvious asymmetry artifacts introduced by LSB replacement can be easily avoided. Therefore, the common approaches used to detect LSB replacement are totally ineffective at detecting the LSBM. Up to now, several steganalytic algorithms (e.g., [7]–[10]) have been proposed to analyze the LSBM scheme. Unlike LSB replacement and LSBM, which deal with the pixel values independently, LSB matching revisited (LSBMR) [1] uses a pair of pixels as an embedding unit, in which the LSB of the first pixel carries one bit of secret message, and the relationship (odd–even combination) of the two pixel values carries another bit of secret message. In such a way, the modification rate of pixels can decrease from 0.5 to 0.375 bits/pixel (bpp) in the case of a maximum embedding rate, meaning fewer changes to the cover image at the same payload compared to LSB replacement and LSBM. It is also shown that such a new scheme can avoid the LSB replacement style asymmetry, and thus it should make the detection slightly more difficult than the LSBM approach based on our experiments. The typical LSB-based approaches, including LSB replacement, LSBM, and LSBMR, deal with each given pixel/pixel pair without considering the difference between the pixel and its neighbors. The pixel-value differencing (PVD)-based scheme (e.g., [17]–[19]) is another kind of edge adaptive scheme, in which the number of embedded bits is determined by the difference between a pixel and its neighbor. The larger the difference, the larger the number of secret bits that can be embedded. Usually, PVD-based approaches can provide a larger embedding capacity. Assuming that a cover image is made up of many no overlapping small sub images (regions) based on a predetermined rule, then different regions usually have different capacities for hiding the message. Generally, the regions located at the sharper edges present more complicated statistical features and are highly dependent on the image contents. In this paper, we propose an edge adaptive scheme and apply it to the LSBMR-based method. The rest of the paper is arranged as follows. Section II analyzes the limitations of the relevant steganography schemes and proposes some strategies. Section III shows the details of data embedding and data extraction in our scheme. Section IV presents experimental results and discussions. Finally, concluding remarks and future.

II. PROPOSED SCHEME

The flow diagram of our proposed scheme is illustrated in Fig. 4. In the data embedding stage, the scheme first initializes some parameters, which are used for subsequent

data preprocessing and region selection, and then estimates the capacity of those selected regions. If the regions are large enough for hiding the given secret message, then data hiding is performed on the selected regions. Finally, it does some post processing to obtain the stego image. Otherwise the scheme needs to revise the Parameters, and then repeats region selection

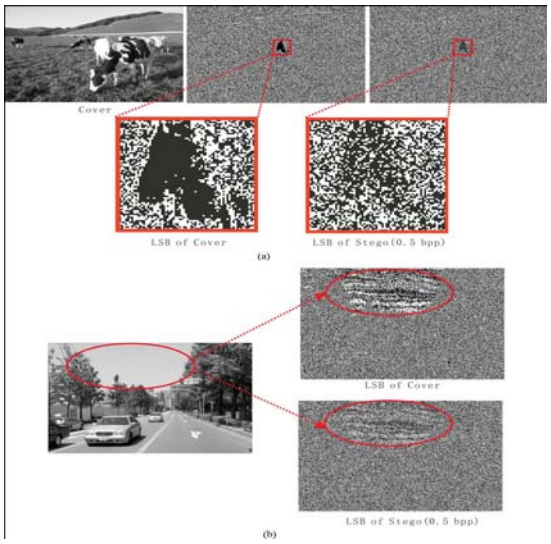


Figure 1.

and capacity estimation until can be embedded completely. Please note that the parameters may be different for different image content and secret message. In data extraction, the scheme first extracts the side information from the stego image. Based on the side information, it then does some preprocessing and identifies the regions that have been used for data hiding. Finally, it obtains the secret message according to the corresponding extraction algorithm. In this paper, we apply such a region adaptive scheme to the spatial LSB domain. We use the absolute difference between two adjacent pixels as the criterion for region selection, and use LSBMR as the data hiding algorithm. The details of the data embedding and data extraction algorithms are as follows.

A. Data Embedding

Step 1: The cover image of size of is first divided into non overlapping blocks of pixels. For each small block, we rotate it by a random degree in the range of, as determined by a secret key. The resulting image is rearranged as a row vector by raster scanning. And then the vector is divided into non overlapping embedding units with every two consecutive pixels, where, assuming is an even number. Two benefits can be obtained by the random rotation. First, it can prevent the detector from getting the correct embedding units without the rotation key, and thus security is improved. Furthermore, both horizontal and vertical edges (pixel pairs) within the cover image can be used for data hiding.

Step 2: According to the scheme of LSBMR, 2 secret bits can be embedded into each embedding unit. Therefore, for a given secret message, the threshold for region selection can be determined as follows. Let be the set of pixel pairs whose absolute differences are greater than or equal to a parameter t

$$EU(t) = \{(x_i, x_{i+1}) \mid |x_i - x_{i+1}| \geq t, \forall (x_i, x_{i+1}) \in V\}$$

Then we calculate the threshold T by

$$T = \arg^{max}_{i}\{2 \times |EU(t)| \geq |M|\}$$

where, is the size of the secret message , and denotes the total number of elements in the set of .

Step 3: Performing data hiding on the set of

$$\mathbf{EU(T)}=\{(x_i, x_{i+1}) \mid |x_i - x_{i+1}| \geq T, \forall (x_i, x_{i+1}) \in V$$

We deal with the above embedding units in a pseudorandom order determined by a secret key . For each unit , we perform the data hiding according to the following four cases.

Case #1:

$$\mathbf{LSB}(x_i) = m_i \,\&\, f(x_i, x_{i+1}) = m_{i+1}$$
$$(x_i', x_{i+1}') = (x_i, x_{i+1});$$

Case #2:

$$\mathbf{LSB}(x_i) = m_i \,\&\, f(x_i, x_{i+1}) \neq m_{i+1}$$
$$(x_i', x_{i+1}') = (x_i, x_{i+1} + \mathbf{r});$$

Case#3:

$$\mathbf{LSB}(x_i) \neq m_i \,\&\, f(x_i - 1, x_{i+1}) = m_{i+1}$$
$$(x_i', x_{i+1}') = (x_i - 1, x_{i+1});$$

Case # 4:

$$\mathbf{LSB}(x_i) \neq m_i \,\&\, f(x_i - 1, x_{i+1}) \neq m_{i+1}$$
$$(x_i', x_{i+1}') = (x_i + 1, x_{i+1});$$

where and denote two secret bits to be embedded. The function is defined as . is a random value in and denotes thepixel pair after data hiding. After the above modifications, and may be out of , or the new difference may be less than the threshold . In such cases,1 we need to readjust them as (x_i'' , x_{i+1}'') **by** (x_i'' , x_{i+1}'') **=arg** $min_{(e1,e2)}\{|e_1 - x_i| + |e_2 - x_{i+1}| | e_1 =$ $x_i' + 4k_1', e_2 = x_{i+1}'$ **+2** $k_2, |e_1 - e_2| \geq T, 0 \leq e_1, e_2 \leq 255, 0 \leq T \leq 31, k_1, k_2 \in Z\}. (*)$

Finally, we have

$$LSB(x_i'') = m_i, f((x_i'', x_{i+1}'') = m_{i+1}$$

Step 4: After data hiding, the resulting image is divided into non overlapping blocks. The blocks are then rotated by a random number of degrees based on. The process is very similar to **Step 1** except that the random degrees are opposite. Then we embed the two parameters into a preset region which has not been used for data hiding. The first one is the block size for block dividing in data preprocessing; another is the threshold for embedding region selection. In all, only 7 bits of side information are needed for each image.

III. CONCLUDING REMARKS

In this paper, an edge adaptive image steganographic scheme in the spatial LSB domain is studied. As pointed out in Section II, there usually exist some smooth regions in natural images, which would cause the LSB of cover images not to

be completely random or even to contain some texture information just like those in higher bit planes. If embedding a message in these regions, the LSB of stego images becomes more random, and according to our analysis and extensive experiments, it is easier to detect. In most previous steganographic schemes, however, the pixel/pixel-pair selection is mainly determined by a PRNG without considering the relationship between the characteristics of content regions and the size of the secret message to be embedded, which means that those smooth/flat regions will be also contaminated by such a random selection scheme even if there are many available edge regions with good hiding characteristics. To preserve the statistical and visual features in cover images, we have proposed a novel scheme which can first embed the secret message into the sharper edge regions adaptively according to a threshold determined by the size of the secret message and the gradients of the content edges.. Furthermore, it is expected that our adaptive idea can be extended to other steganographic methods such as audio/video steganography in the spatial or frequency domains when the embedding rate is less than the maximal amount.

REFERENCES

[1] J. Mielikainen, "LSB matching revisited," IEEE Signal Process. Lett., vol. 13, no. 5, pp. 285–287, May 2006.
[2] A.Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in Proc. 3rd Int. Workshop on Information Hiding, 1999, vol. 1768, pp. 61–76.
[3] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color, and gray-scale images," IEEE Multimedia, vol. 8, no. 4, pp. 22–28, Oct. 2001.
[4] S. Dumitrescu, X.Wu, and Z.Wang, "Detection of LSB steganography via sample pair analysis," IEEE Trans. Signal Process., vol. 51, no. 7, pp. 1995–2007, Jul. 2003.
[5] A.D. Ker, "A general framework for structural steganalysis of LSB replacement," in Proc. 7th Int. Workshop on Information Hiding, 2005, vol. 3427, pp. 296–311.
[6] D. Ker, "A funsion of maximum likelihood and structural steganalysis," in Proc. 9th Int. Workshop on Information Hiding, 2007, vol. 4567, pp. 204–219.
[7] J. Harmsen and W. Pearlman, "Steganalysis of additive-noise modelable information hiding," Proc. SPIE Electronic Imaging, vol. 5020, pp. 131–142, 2003.
[8] A.D. Ker, "Steganalysis of LSB matching in grayscale images," IEEE Signal Process. Lett., vol. 12, no. 6, pp. 441–444, Jun. 2005.
[9] F. Huang, B. Li, and J. Huang, "Attack LSB matching steganography by counting alteration rate of the number of neighbourhood gray levels," in Proc. IEEE Int. Conf. Image Processing, Oct. 16–19, 2007, vol. 1, pp. 401–404.
[10] X. Li, T. Zeng, and B. Yang, "Detecting LSB matching by applying calibration technique for difference image," in Proc. 10th ACM Workshop on Multimedia and Security, Oxford, U.K., 2008, pp. 133–138.
[11] Y. Q. Shi et al., "Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network," in Proc. IEEE Int. Conf. Multimedia and Expo, Jul. 6–8, 2005, pp. 269–272.
[12] Li, J. Huang, and Y. Q. Shi, "Textural features based universal steganalysis," Proc. SPIE on Security, Forensics, Steganography and Watermarking of Multimedia, vol. 6819, p. 681912, 2008.
[13] M. Goljan, J. Fridrich, and T. Holotyak, "Newblind steganalysis and its implications," Proc. SPIE on Security, Forensics, Steganography and Watermarking of Multimedia, vol. 6072, pp. 1–13, 2006.
[14] K. Hempstalk, "Hiding behind corners: Using edges in images for better Steganography," in Proc. Computing Women's Congress, Hamilton, New Zealand, 2006.
[15] K. M. Singh, L. S. Singh, A. B. Singh, and K. S. Devi, "Hiding secret message in edges of the image," in Proc. Int. Conf. Information and Communication Technology, Mar. 2007, pp. 238–241.
[16] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images," in Proc. IEEE on Digital Signal Processing Workshop, Sep. 1996, pp. 37–40.
[17] Wu and W. Tsai, "A steganographic method for images by pixelvalue differencing," Pattern Recognit. Lett., vol. 24, pp. 1613–1626, 2003.
[18] X. Zhang and S. Wang, "Vulnerability of pixel-value differencing steganography to histogram analysis and modification for enhanced security," Pattern Recognit. Lett., vol. 25, pp. 331–339, 2004.
[19] H. Yang, C. Y. Weng, S. J. Wang, and H. M. Sun, "Adaptive data hiding in edge areas of images with spatial LSB domain systems," IEEE Trans. Inf. Forensics Security, vol. 3, no. 3, pp. 488–497, Sep. 2008.
[20] M. Kharrazi, H. T. Sencar, and N. Memon, "Cover selection for steganographic embedding," in Proc. IEEE Int. Conf. Image Processing, Oct. 8–11, 2006, pp. 117–120.

BIOGRAPHY

Simi Sara Mani graduated from Karunya University, in information technology during the year 2008. She obtained her Master degree in Computer Science and Engineering from Faculty of Engineering, Avinashlingam Deemed University for Women, Coimbatore in the year 2011. At present she is a lecturer in the Department of Computer Science and Engineering, Easa College of engineering & technology, Coimbatore, India. Her area of interest includes Data mining and Web Services. She has 1 year of experience in teaching.

S. Sivaranjani graduated from Anna University, in Computer Science and Engineering during the year 2005. She obtained her Master degree in Computer Science and Engineering from Anna University of Technology, Coimbatore in the year 2010. At present she is an assistant professor in the Department of Computer Science and Engineering, Faculty of Engineering, Avinashlingam Deemed University for Women, Coimbatore, India. Her area of interest includes Data mining and Web Services. She has 5 years of experience in teaching.

Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

Securing Wireless Bluetooth Sensor Systems

[1] Prof. Anand Nayyar

***Abstract*— In this research paper a Low power, Portable and Secure Wireless Bluetooth Sensor System has been designed and its performance has been evaluated. The sensor system is light weight and has interoperability with Personal Area Network (PAN) and the architecture has been implemented by adopting an FPGA and a Bluetooth Module. The analysis of design shows its capability of continuous transmission of analog signals and a high rate of security level. As low sampling rates, the adopted solution offers low power consumption and lower battery capacity can be adopted and sensor weight can be minimized. With higher sampling rates, the Wireless Sensor System is equipped with FGMA which offers best architecture solution and high performance. So Wireless Bluetooth Sensor system can be widely adopted in critical applications like Detecting Vital Signs in Patients having serious pathologies.**

***Index Terms*— Bluetooth, Wireless sensor systems, Wireless Technology, Security, IEEE 802.15.1**

I. INTRODUCTION

Bluetooth, A technology which requires no introduction and is being used constantly in Mobile Phones, Computers, Tablet PC'S, TV'S, Gaming Consoles and much more for transferring the data from one device to another. Bluetooth Wireless technology is becoming very popular to replace existing short range wired technology with short-range Wireless technology to enable new types of applications.

With the increase in use of Bluetooth Technology, Various Researches and Manufactures are closely working to use this technology in completely different environments such as in Medical sector to improve the life quality and to reduce the cost incurred by hospitals in treating patients.

A new concept called PAN (Personal Area Network) is evolving along with Bluetooth Technology. A PAN consists of a limited number of units interconnected to form a network and to exchange information among the connected nodes. Bluetooth acts a local connection interface between different personal units like Mobile Phones, PDA's, Keyboard, Mouse, Gaming Consoles and much more. Bluetooth is a true enabling technology for the PAN Vision. The units are typically consumer devices which are used by different manufactures in different ways. So in order to have better interoperability between the personal devices, the security level has to be set up by the user. The Bluetooth technology has been designed in such a intelligent manner which enables even a ordinary user to maintain a good security level to protect the data and communication links in operation.

With the help of PAN Technology, users can access their data wirelessly between different devices, work on them and store them in an Information System or in Electronic Personal Record.

Wireless Systems implementation has been done by taking into account the following important key points:-

1. Using Wireless devices everywhere and avoiding the location lockdown
2. Interoperability with other technologies for Communications.
3. Enough security to prevent eavesdropping and intrusion avoidance.
4. System high level interface immunity.

But if we implement any technology we have certain expectations. Same is the case with Bluetooth Technology. The following are the desired expectations from the Bluetooth Technology:-

1. Confidentiality Protection.
2. Only authenticated devices should communicate in the Bluetooth Network.
3. Easy to use as well as High Speed Data Connectivity.
4. Proper Security Measures to avoid Malicious Activity and DoS Attacks.

In this Research Paper, A Sensor System Architecture is presented along with its benefits. As well as the Bluetooth technology security is also analysed and solution is proposed.

II. BLUETOOTH STANDARD

Bluetooth is a proprietary open wireless standard which was created by Telecoms vendor Erricson in 1994 for exchanging data over short distances using short wavelength radio transmissions from fixed and mobile devices creating Personal Area Networks (PANs) with higher degree of security. Basically Bluetooth standard was designed to replace RS-232 data cables. Bluetooth is regarded as developing network technology which is able to support data nd voice communications and is characterized by low complexity, robustness and low power cum cost.

Bluetooth uses a Radio Technology called FHSS (Frequency Hopping Spread Spectrum) which chops up the data being sent and transmits the chunks of data on 79 bands (1 MHz each) in the range of 2402-2480 MHz. This range is in the globally, unlicensed Industrial, Scientific and Medical (ISM) 4.4 GHz short-range radio frequency band.

Bluetooth has the ability to form PANs and is regarded as Packet-Based Protocol with a Master-Slave structure. One master can communicate with up to 7 slaves in a piconet; all the devices share the master's clock. When a device is present simultaneously in more than one piconet, a scatternet is established. The master establishes the hop sequence and communicates with active slaves using TDM (Time Division Multiplexing) Technique in which the time is divided into 625 μs intervals called slots. The transmission between master and slave starts in even the numbered slots while the slave to master transmission starts in odd numbered slots. Master and slaves are allowed to transmit the packets in 1, 3,5 consecutive slots. Forward Error Correction (FEC), Cyclic

Manuscript received Mar 10, 2011.

Prof. Anand Nayyar, P.G. Department of Computer Science, Kamla Lohtia Sanatam Dharam College, Ludhiana, (e-mail : anand_nayyar@yahoo.co.in)

Redundancy Check (CRC), Header Error Check (HEC) and Automatic Reepeat Request (ARQ) are the techniques whihc provide data protection against imperfect channels.

The Packet Exchange is based on the basic clock, defined by the master, which ticks at 312.5 µs intervals. Two clock ticks make up a slot of 625 µs; two slots make up a slot pair of 1250 µs. Compared with other systems in the same fry band, the Bluetooth Radio hops is very faster and uses shorter packets. There are 79 channels 1 MHz bandwidth, starting from 2.402 GHz to 2.480 GHz.

The Bluetooth Technology provides high security mechanisms including a globally unique six byte Bluetooth Device Address (BDA), authentication, authorization, encryption and PIN exchange at user level. In general, Bluetooth Security is divided into three modes: (a) Non-Secure; (b) Service level enforced security (c) Link level enforced security. In non-secure, a Bluetooth device doesn't initiate any secure measures. In service-level enforced security mode, "two Bluetooth devices can establish a non secure Asynchronous Connection-Less (ACL) Link. In the link level enforced security, the Bluetooth device initiates securit procedures before the channel is established.

III. BLUETOOTH SENSOR ARCHITECTURE

The Bluetooth Sensor Architecture consists of seven client modules and one master module for System Control. The main component of Bluetooth Architecture is FPGA which is shown in diagram 1. FPGA allows the device to be programmed, debugged and reconfigured after it is soldered onto a printed circuit board which reduces the possibility of lead damage and electrostatic discharge exposures.

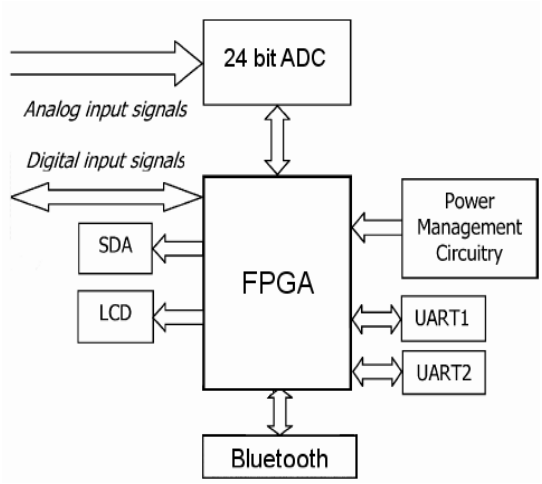


Figure.1 FPGA-The Main Component of Module

In this research paper, signals are generated by biomedical sensors for monitoring critical parameters such as Vital signs in patients. It has been realized in Wireless Sensor Architecture using one Analog/Digital Convertor (ADC) and two processors sharing the Bluetooth stack. A 24-bit multiplex sigma-delta converter converts the analogue input signal with 0-5Volt range. The sampling rate is 500 Hz on each of two channels. The digital signals are transmitted to a remote acquisition master sensor via Bluetooth (PAN 1540) .The FPGA controls the acquisition from the sigma-delta converter and, as soon as an AD conversion has been made, saves that particular value in the FPGA internal RAM memory.

The FPGA sends the data from the memory to the Bluetooth module while controlling and storing the new ADC value. The Bluetooth™ management is implemented in the FPGA and controls the Bluetooth™ module. The FPGA construct and decodes the Host Controller Interface (HCI) packages in order to establish connections and manage data communications. The communications between the FPGA and Bluetooth™ is done by serial UART as shown in Diagram 2.

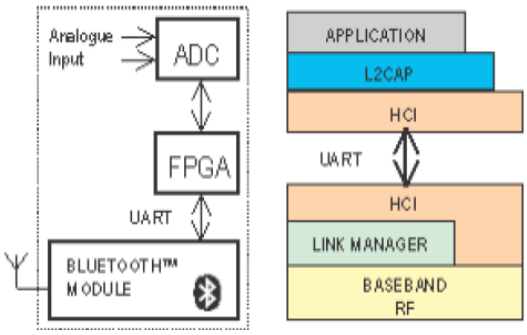


Figure.2 FPGA Control Management

The power management circuit, which powers all the modules, consists of linear voltage regulators to provide positive and negative voltages from a PP3 9-V battery with a rating of 550 mAh. The regulators have a maximum current drain of approximately 500 mA, which although high still allows over an hour of continuous operation. In idle state, the current drain is less than 1.5 mA.

The minimal solution with only 1-chip wireless sensor using the internal uncommitted 8-bit ADC of the PAN1540 is possible. This implementation is an embedded solution where the Bluetooth™ module executes a Virtual Machine (VM) application.

Pan1540 has three general purpose analog interface pins; two of them are used as analog inputs for the ADC, which acts as input channel for a sensor signal.

The ADC is controlled by user code, which is interpreted by the VM when the scheduler runs the task.

This solution has been revealed unsatisfactory because the PAN 1540 allows only a limited number of instructions of the VM before changing context. Therefore, there is no guarantee that the ADC will be controlled in real time while another process starts. Moreover, PAN1540 does not support a Real-Time Operating System (RTOS) because the execution latency of embedded code is random.

IV. BLUETOOTH SECURITY SENSOR SYSTEM NETWORKING

It is very difficult to find the correct level of security when a new communication technology evolves. It is also very difficult in the case of Bluetooth. In order to offer interoperability and to provide support for a specific application, it has developed a set of profiles. A profile is an umambigious description of communication interface between two units for one particular service.

As the main purpose behind the development of Bluetooth technology is to replace short range cables. Pure cable

replacement is done through RS232 emulation which is offered by serial port profile. Several other profiles like PAN (Personal Area Network) and LPP (Local Positioning Profile) make use of serial port profile.

This section primarily discusses the security issues and solutions for remote access to SIM (Subscription Identity Module) over Bluetooth connection. A SIM card is regarded as an Integrated Circuit used in GSM Mobile Telephony to store subscriber information. In this research paper a SIM Solution is implemented inside the FPGA ROM memory. Altera QuartusII v5.1 software generates FPGA configuration data file which is stored in FLAH ROM of processor and memory module.

This SIM information is used to connect a remote sensor to a master network in a secured manner (Laptop, Smartphone, PDA Device or Tablet Device) which makes it possible for mobile network operator to identify the subscribers using the network as well as it also allows the operator to enable the connect of mobile network services. The Bluetooth SIM Access Profile defines protocols and procedures for the access to a remote SIM over Bluetooth Serial Port (RFCOMM) Connection. The SIM Access Profile Communication Stack is defined in Diagram 3. The SIM access messages consist of a header and payload. The header describes the type and the number of parameters transferred in the message. Messages have been defined for the remote control of the SIM sensor and for transfer SIM messages.

Two different roles are defined in the profile:

1. SIM ACCESS CLIENT
2. SIM ACCESS SERVER

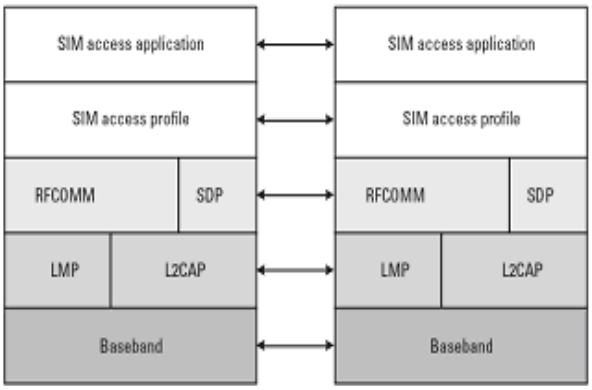


Figure.3 Sim Access Profile Communication Stack

The SIM access client uses the SIM access profile for the connection to another device, the SIM access server, over the Bluetooth. The adopted interconnectivity system is defined in Diagram 4.

In this scenario, seven SIM access clients are wireless interconnected with one SIM access server (laptop) within PAN wireless network. A SIM access is needed for the subscriber authentication inside the wireless network. The laptop has an integrated Bluetooth module and uses the SIM access profile to access it.

In the implemented sensor architecture, the SIM is used for security critical services in security mode 3 with a 32-digits pass-key.

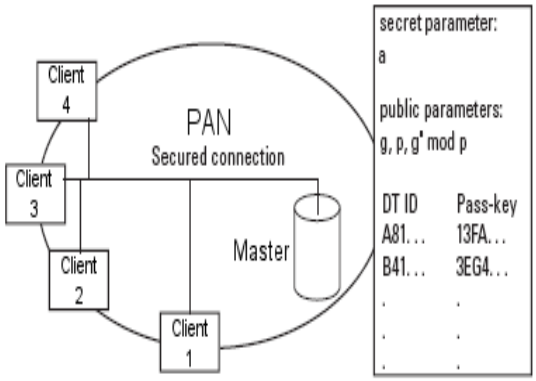


Figure.4 Pan Secured Interconnection System

In the FPGA ROM, a 128-bits encryption key has been implemented for a major security level. To avoid the typing of the 32 digits pass-key by the user, in this system the pass-key value is generated by the server and displayed to the user. The security required by the SIM access profile gives the necessary protection for the message exchange between the client and the server. However, to avoid security holes in the master SIM access server implementation, additional security measures has been developed in the implemented architecture.

One problem is that in an implementation that just follows the specification, all the messages from the client to the server have to be accepted and forwarded to the SIM. This is a potential security risk for the sensitive functions in the subscription module, available for the remote device. This device might have been compromised in some way or it might have been infected by a virus or other harmful software. For this reason, the access to the subscription module by the server has to be restricted.

This can be achieved if, at the security pairing, the server selects the set of services in the SIM that the client should be allowed to access. Then the record of allowed services has to be stored in a special and protected access control database. When the client has been authenticated against the server, a filtering process or a security filter has to check all messages from the client to the subscription module, as is illustrated in Diagram 5

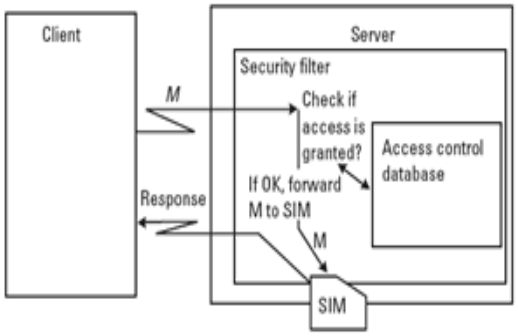


Figure.5 Sim Client- Server Access Control

The filter makes sure that only messages allowed according to the access database are forwarded to the subscription module.

V. CONCLUSION

In this research paper Bluetooth System has been designed and its performance has been evaluated on security parameters. The solution proposed will reduce the number of components and also the power consumption allowing longer battery lifetime. But of Future development, the ZIGBEE standard will be considered to optimize the power consumption performance of the remote monitoring system.
In order to increase the level of network security, A SIM solution is proposed in security mode 3 with 32 digits pass-key. The security required by SIM Access Profile gives the necessary protection for message exchange between client and server.

REFERENCES

[1] Brooks, T, "Wireless technology for industrial sensor and control networks" Sensor for Industry, 2001, Proceedings of the First ISMEEE Conference, Page(s):73 -77, 2001.

[2] Rauchhaupt, L.; "System and Device Architecture of a Radio Based Fieldbus -The Rfieldbus System" IEEE International Workshop on Factory Coomunication systems. Page@): 185-192, Aug 2002.

[3] G.J. Pottie and W.J. Kaiser, "Wireless Integrated Network Sensors", Commun. ACV, vol43, pp. 51.58, no 5 May 2000

[4] R. S. H. Istepanian, "Modeling of GSM-based mobile telemedical system," in Proc. 20th Annu. IEEE/EMBS Conf., vol. 20, Hong Kong, 1998, pp. 1166–1169.

[5] Kansal, A.; Desai, U.B.; "Bluetooth primer" Internetdocument,http://www.ee.ucla.edu/kansal/bt_primer.pdf Page: 4, 2002.

[6] Baatz, S.; Frank, M.; Gopffarth, R.; Kassatkine, D.; Martini, P.; Schetelig, M.; Vilavaara, A.; "Handoff support for mobility with IP over Bluetooth" Local Computer Networks, 2000.Proceedings. 25th Annual IEEE Conference on, Page(s): 143 -154, 2000.

[7] Bluetooth TM SIG; "Specification of the BluetoothTM System Core 1.11'' http://www.bluetooth.com, Vol: 1 , Page: 65,2001.

[8] J. S. Park and D. Dicoi, "WLAN security: Current and future," IEEE Internet Comput. vol. 7, no. 5, pp. 60–65, Sep./Oct. 2003.

[9] J. Andreasson, J. G. Castaño, M. Lindén, Y. Bäcklund, "Remote System for Patient Monitoring Using Bluetooth™". Proc. 2nd International Symposium on Telemedicine, Gothenburg, Sweden, 2002.

[10] J. Andreasson, M. Ekstrom, A. Fard, J. G. Castano, T. Johnson, "Remote system for patient monitoring using Bluetooth". Prec. 1st IEEE int. conf on Sensors, Orlando, USA, 2002 pp 304-307

[11] J. G. Castaño, J. Lönnblad, M. Svensson, A. G. Castaño, M. Ekström and Y. Bäcklund, "Steps towards a Minimal Mobile Wireless Bluetooth™ Sensor" Proc. 2004 Sicon, New Orleans, USA, 2004 pp 79-84.

[12] Internet document www.panasonic-eutc.com /products / daten/pdf/Web_PAN1540-C.pdf

[13] 3rd Generation Partnership Programme, 3GPP TS 11.11, Specification of the Subscriber Identity Module Mobile Equipment (SIM-ME) Interface, Version 8.10.0, September 2003.

[14] D. A. Bonnett, "Design for in-system programming," in Proc. Int. Test Conf., Atlantic City, NJ, 1999, pp. 252–259.

[15] M. Winters, "Using IEEE-1149.1 for in-circuit emulation," in WESCON/94 Idea /Microelectronics Conf. Rec., 1994, pp. 525–528.

[16] J. Andrews, "An embedded JTAG, system test architecture," in Proc. Electro/94 Int. Conf., Boston, MA, 1994, pp. 691–695

[17] M. Bogdan, H. Sanders, M. Shochet, and A. Amadon, "Dual method of configuring Altera 10 K family PLDs," in Proc. 11th IEEE NPSS Real Time Conf., Santa Fe, NM, 1999, pp. 312–314.

[18] "Using the Jam Language for ISP & ICR via an Embedded Processor," Altera Corp., San Jose, CA, Altera Application Note 88, Version 3.01, Nov. 1998.

BIOGRAPHY

Prof. Anand Nayyar (B.Com, M.C.A, M.Phil, M.Tech) currently working as Assistant Professor in P.G. Department of Computer Science in Kamla Lohtia Sanatam Dharam College (KLSD College), Ludhiana. The Author possess many International Credentials like A+,CCNA, MCSE, MCTS, MCITP, RHCE, OCP, CEH, MCSD.net to name a few. The author has published AROUND 35 research papers in National and 8 research papers in International Conferences and published 6 books on topics like Networking, Database, Data Structures and Information Technology Fundamentals. His areas of interest include Networking, Distributed Systems, Linux & Open Source Technology, Database Management Systems, Software Engineering and Testing, Computer Graphics, Information Systems and Digital Image Processing.

The Author is permanent member of research organizations like IAENG (International Association of Engineers), IACSIT (International Association of Computer Science and Information Technology).

Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

A Survey on Applications of Fuzzy Logic in Agriculture

[1] Sanjeev S Sannakki, [2] Vijay S Rajpurohit [3] Arunkumar.R

Abstract— **The sole area that serves the food needs of the entire human race is the Agriculture sector. Research in agriculture is aimed towards increase of productivity and food quality at reduced expenditure and with increased profit. The challenge of the precision approach is to equip the farmer with adequate and affordable information and control technology. Methods for identification of diseases found in any parts of the plant play a critical role in disease management. Consequently minimizing plant diseases allows substantially improving quality of the products. Many methods and techniques of image processing and soft computing are applied on a number of plants for early detection and diagnosis of different plant diseases. Since fuzzy logic can effectively handle the vague image data, present paper discusses several aspects and techniques of precision agriculture which employ Fuzzy logic.**

Index Terms— **Fuzzy Logic, Support Vector Machine (SVM).**

I. INTRODUCTION

The objectives of precision agriculture are profit maximization, agricultural input rationalization and environmental damage reduction, by adjusting the agricultural practices to the site demands. Plant disease is one of the crucial causes that reduces quantity and degrades quality of the agricultural products. The ability of disease diagnosis in earlier stage is very important task. There are numerous characteristics and behaviors of such plant diseases in which many of them are merely distinguishable. Hence an intelligent decision support system for Prevention and Control of plant diseases is needed which is an integrated agricultural information platform, that uses some high-tech and practical technology, such as fuzzy logic, neural networks, support vector machines and such other soft computing techniques to appropriately detect and diagnose the plant diseases.

Fuzzy set theory is an extension of conventional set theory that deals with the concept of partial truth. Fuzzy logic aims to model the vagueness and ambiguity in complex systems. In many image processing applications, expert knowledge must be used for applications such as object recognition and scene analysis. Fuzzy set theory and fuzzy logic provide powerful tools to represent and process human knowledge in the form of fuzzy IF-THEN rules.

Manuscript received Mar 11, 2011.

Sanjeev S Sannakki,, Department of CSE,Gogte Institute of Technology, Belgaum, Karnataka, India. (e-mail :sannakkisanjeev@yahoo.co.in)
Vijay S Rajpurohit, Department of CSE, Gogte Institute of Technology, Belgaum, Karnataka, India. (e-mail : vijaysr2k@yahoo.com)
Arunkumar.R, Department of CSE, Gogte Institute of Technology, Belgaum, Karnataka, India. (e-mail : kumararun37@gmail.com)

II. RELATED WORK

Over the past few decades, fuzzy logic has been used in a wide range of problem domains. The areas of applications are very wide: process control, management and decision making, operations research, economies and pattern recognition and classification. In the lack of precise mathematical model which will describe behavior of the system, Fuzzy Logic is a good "weapon" to solve the problem: it allows using logic if-then rules to describe the system's behavior.

In the paper of "Image Classification Based on Fuzzy Logic" a prior knowledge about spectral information for certain land cover classes is used in order to classify SPOT image in fuzzy logic manner. More specifically, input (image channels) and output variables (land classes) are introduced in Matlab's environment, membership functions are defined using results from supervised classification which was conducted with PCI ImageWorks®, Matlab's Fuzzy Logic Toolbox was then used in definition of fuzzy logic inference rules, these rules are tested and verified through the simulation of classification procedure at random sample areas and at the end, SPOT image classification was conducted.

Output images coming from PCI maximum likelihood (ML) and fuzzy classification can be compared which is shown in fig 1. These grayscale images are produced in such a way that pixels coming from the same class have the same digital numbers in both images: water (50), urban (100), crop 1 (150), crop 2 (200) and vegetation (250). This is the basis for image comparison. Percentage of classified pixels in both methods is given in the following table:

TABLE I. PERCENTAGE OF CLASSIFIED PIXELS IN ML AND FUZZY CLASSIFICATION

Method class	PCI	Fuzzy	Difference
Water	1.25	1.39	0.14
Urban	15.62	13.95	1.67
Crop1	13.1	17.24	4.14
Crop2	28.82	34.11	5.29
Vegetation	37.90	29.99	7.91

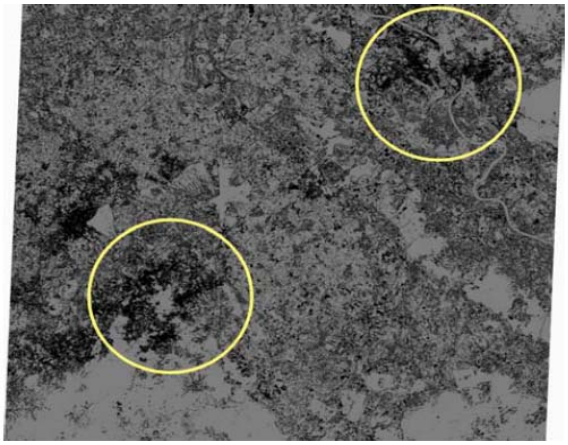


Figure.1 ML and Fuzzy classification comparison image Courtesy: Reference [1]

The experimental results showed that fuzzy logic can be satisfactorily used for image classification providing a greater level of classification accuracy.

The main objective presented in the paper "Recognition of Weeds with Image Processing and Their Use with Fuzzy Logic for Precision Farming" is to develop a methodology for processing digital images taken from cornfields in order to determine a weed map. Based on the weed map, a program was then developed to simulate the control of an herbicide sprayer. Given that information concerning economic thresholds of weed impact on crop productivity cannot easily be adapted to a given region or even to a given farm, the researchers decided that the fuzzy logic approach should be employed to convert image data into sprayer command and the existing fuzzy logic controller was limited to the control of one nozzle. The Fuzzy Logic Toolbox v2.0 of MATLAB was used to develop the fuzzy logic model for the herbicide application. In this project, a fuzzy logic system was developed to simulate human decision-making in determining herbicide application based on greenness and patch size. There are three components in a fuzzy logic system: fuzzy values for inputs and outputs, a set of fuzzy rules, and fuzzy inference mechanism. In fuzzy inference, several fuzzy membership functions are developed to generate a degree of truth. The fuzzy logic herbicide application model was tested on a hypothetical field to determine the potential herbicide savings. The reductions in herbicide use compared to a uniform application for different combinations of weed coverage and weed patch thresholds are listed in the following table:

TABLE II. RESULTS FOR REDUCTION IN HERBICIDE USE (%) BY VARIABLE-RATE APPLICATION USING DIFFERENT THRESHOLD VALUES

Weed Patch Threshold	Weed coverage threshold				
	1%	2%	3%	4%	5%
1%	4.86	6.44	8.28	10.33	12.76
2%	5.41	7.32	9.34	11.52	13.91
3%	6.39	8.74	11.38	14.19	16.71
4%	7.80	10.60	13.93	17.43	20.32
5%	9.78	12.82	16.66	20.72	24.03
On/Off Application	14.56	24.97	39.69	53.57	63.72

The results of this study have shown that weeds can be located by the greenness method and a fuzzy logic controller automatically manages herbicide applications to obtain effective weed control, reduce costs, and minimize soil and water pollution.

In "Agricultural Produce Sorting and Grading using Support Vector Machines and Fuzzy Logic" an automated grading system has been proposed and designed to overcome the problems of manual grading. It combined three processes – feature extraction, sorting and grading without any human intervention. Initially the images were taken using a regular digital camera. The feature extraction process was done using the MATLAB image processing toolbox. Following is the resultant image with extracted features.

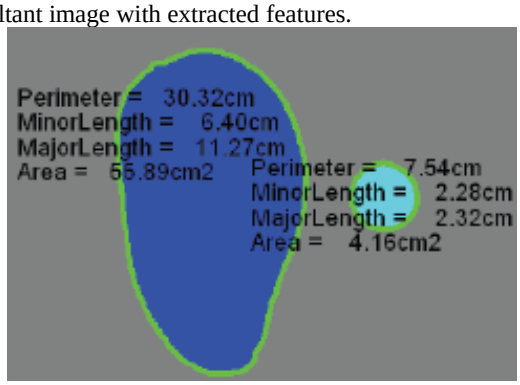


Figure.2 Image with extracted features Courtesy: Reference [3]

Then shape sorting was done using the SVMs that has the ability to recognize the shape of an object. The sorting task involves two sets of data; i.e., training data and validation data. Each instance of training data consists of one 'target value' (class) and several features. The most important role in shape recognition is feature extraction. The features are normalized by

$$\text{Xnorm} = \frac{Xraw - X\min}{X\max - X\min} \quad (1)$$

where Xraw, Xnorm, Xmax, Xmin are the raw, normalized, minimum and maximum values of the features. This ensures that Xnorm lies in (0, 1), which is very important for SVM classification.

Table III shows the classification accuracies of five fruits:

TABLE III CLASSIFICATION RESULT

Specimen	Classificatio

	n Accuracy
Apple	96.25%
Banana	81.25%
Carrot	0%
Mango	98.75%
Orange	6.25%

Fuzzy Logic was then applied for the agriculture produce grading. This technique was chosen because it represents a good approach when human experience needs to be incorporated into the decision making process. The grade is determined based on fruit type and fruit features. Following figure shows the Fuzzy grading system that uses 3 inputs to determine the output (size) of the fruit: major length, minor length and area.

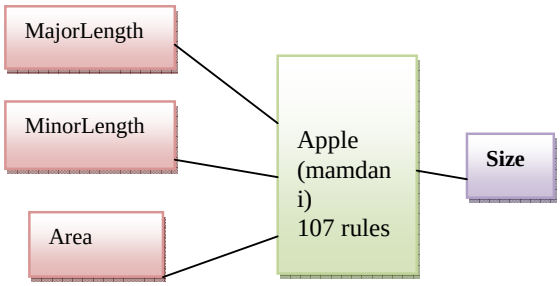


Figure.3 Fuzzy Grading System

Following table shows the grading result of the fruits, which represents very good grading accuracy:

TABLE IV. GRADING RESULT

Specimen	**Grading Accuracy**
Apple	98.85%
Banana	98.75%
Mango	89.74%

III. CONCLUSION

So far we have discussed various image processing techniques, which employ fuzzy techniques and inference rules, and their role in wide range of precision agriculture applications such as feature extraction, texture analysis, agriculture produce grading, effective use of herbicide sprayers in disease control etc. Reference [1] infers that considering the level of classification accuracy, fuzzy logic can be satisfactorily used for image classification. It is recommended to use a fuzzy logic controller for effective weed control as proved in reference [2]. In a new technique for sorting and grading [3] fuzzy logic has been employed for grading whose results were proven to be good for three of the five chosen fruits. In all the research papers discussed above, authors have shown that fuzzy based approaches outperformed comparatively.

IV. ACKNOWLEDGEMENT

We thank Visvesvaraya Technological University, Belgaum, India for funding this work. We express our sincere thanks to Dr. V. B. Nargund, plant pathologist, University of Agricultural Sciences, Dharwad, India, for his kind help on plant pathology.

V. REFERENCES

[1] I. Nedeljkovic, “Image Classification Based On Fuzzy Logic”, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34,

[2] C.C. Yang1, S.O. Prasher, J.-A. Landry, J. Perret1 And H.S. Ramaswamy , “Recognition OfWeeds With Image Processing And Their Use With Fuzzy Logic For Precision Farming”, Canadian Agricultural Engineering Vol. 42, No. 4

[3] Nur Badariah Ahmad Mustafa, Syed Khaleel Ahmed, Zaipatimah Ali, Wong Bing Yit, Aidil Azwin Zainul Abidin, Zainul Abidin Md Sharrif, “Agricultural Produce Sorting and Grading using Support Vector Machines and Fuzzy Logic”, 2009 IEEE International Conference on Signal and Image Processing Applications

[4] Mario I. Chacon, Luis Aguilar, Abdi Delgado, “Definition And Applications Of A Fuzzy Image Procesing Scheme”, 0-7803-81 16-5/02/ 2002 IEEE.

[5] Patricia Melin, “Interval Type-2 Fuzzy Logic Applications in Image Processing and Pattern Recognition”, 2010 IEEE International Conference on Granular Computing

[6] Ching-Yu Tyan and Paul P. Wang, “Image Processing - Enhancement, Filtering and Edge Detection Using the Fuzzy Logic Approach”, 0-~803-0614-7/93 1993EEE

[7] E.G. Dunn', J.M. Kellet, L.A. Marks', J.E. Ikerd', P.D. Gadet, and L.D. Godsey', “Extending the Application (of Fuzzy Sets to the Problem of Agricultural1 Sustainability”, 0-8186-7126-2/95 1995 IEEE Proceedings of ISUMA-NAFIPS '95

[8] Murali Siddaiah Michael A. Lieberman Nadipuram R. Prasad, “Identification of Trash Types in Ginned Cotton using Neuro Fuzzy Techniques”, 1999 IEEE International Fuzzy Systems Conference Proceedings August 22-25, 1999, Seoul, Korea

[9] Nur Badariah Ahmad Mustafa, Nurashikin Ahmad Fuad, Syed Khaleel Ahmed, Aidil Azwin Zainul Abidin, Zaipatimah Ali, Wong Bing Yit, and Zainul Abidin Md Sharrif, “Image Processing of an Agriculture Produce: Determination of Size and Ripeness of a Banana”, 978-1-4244-2328-6/08/ © 2008 IEEE

[10] G. Louverdis and I. Andreadis, “Design and Implementation of a Fuzzy Hardware Structure for Morphological Color Image Processing”, IEEE transactions on circuits and systems for video technology, vol. 13, no. 3, march 2003

[11] Tian You-Wen, WANG Xiao-Juan, ” Analysis of Leaf Parameters Measurement of Cucumber Based on Image Processing”, World Congress on Software Engineering, 978-0-7695-3570-8/09 © 2009 IEEE

[12] A.Meunkaewjinda, P.Kumsawat, K.Attakitmongcol and A.Srikaew, “Grape leaf disease detection from color

imagery using hybrid intelligent system", Proceedings of ECTI-CON2008 978-1-4244-2101-5/08/ ©2008 IEE

[13] Qing Yao, Zexin Guan, Yingfeng Zhou Jian Tang, Yang Hu, Baojun Yang, "Application of support vector machine for detecting rice diseases using shape and color texture features", International conference on Engineering Computation, 2009

[14] A.Camargo, J.S.Smith, "Image pattern classification for the identification of disease causing agents in plants", Computers and Electronics in Agriculture 66 (2009), 121-125

[15] Narendra V G, Hareesh K S, "Quality Inspection and Grading of Agricultural and Food Products by Computer Vision- A Review", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.1, May 2010

[16] A.Camargo, J.S.Smith, "An image-processing based algorithm to automatically identify plant disease visual symtoms", Biosystems Engineering I02 (2009), 9-21

BIOGRAPHY

Sanjeev S Sannakki, Research Scholar, Visvesvaraya Technological University (VTU), has completed his post graduation with the specialization in Computer Networking. His career spans over a decade in the field of teaching, research and other diversified in-depth experience in academics. He is currently working as Assistant Professor in the Department of CSE, Gogte Institute of Technology, Belgaum. He has published various papers in national/international conferences and journals. He is also guiding the projects of UG/PG students of VTU.

Vijay S Rajpurohit is working as Professor, in the department of Computer Science and Engg at Gogte Institute of Technology, Belgaum. He did his B.E. in 1997, M.Tech at N.I.T.K Surathkal in 2003 and PhD from Manipal University, Manipal in 2009. He has presented papers in various national/international conferences and published papers in international journals. He is the reviewer for three international conferences and 05 international journals. Currently he is carrying out research under the grant of Visveswaraya Technological University Belgaum,India, in the Agricultural domain.

Arun Kumar R secured B.E. Degree in Computer Science & Engineering stream from Bapuji Institute of Engineering and Technology, Davangere in the year 2009 and pursuing M.Tech degree in Computer Science & Engineering stream in Gogte Institute of Technology, Belgaum under Visvesvaraya Technological University. His areas of interest include image processing and machine learning.

Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

Advanced Fault Detection Scheme for AES Architecture

[1]S.Anandi Reddy

***Abstract*— Cryptography is a method that has been developed to ensure the secrecy of messages and transfer data securely. The Advanced Encryption Standard (AES) is the newly accepted symmetric cryptography standard for transferring block of data securely. However, the natural and malicious injected faults reduce its reliability and may cause confidential information leakage. The objective of this paper is to find optimized fault detection schemes for reaching reasonable fault coverage in the high performance AES implementations. In order to provide low cost complexity signature, two sets of error indication flag is used. Thisstructure can be applied to both look-up tables and logic gate for the implementation of S-box and inverse S-box and their parity predictions. Defects in the logic gates causedeither by the natural faults or malicious injected faults that are detected independent of the method the S-box is implemented Moreover, the overhead costs, including space complexity and time delay of the proposed schemes are analyzed. Finally, our simulation results show the error coverage of greater than 99 percent for the proposed schemes.**

***Index Terms*— Advanced Encryption Standard, S-Box, inverse S-box, composite field, fault detection.**

I. INTRODUCTION

1.1 SYMMETRIC KEY CRYPTOGRAPHY

Cryptography is a method that has been developed to ensure the secrecy of messages and transfer data securely. In digital communications the data is sent through the wires or air and thus it is not protected from eavesdropping. Therefore, confidentiality of the transferring data is of extreme importance. Encryption is a process which transforms the data that is aimed to be sent to an encrypted data using a key. The encryption process is not confidential but the key is only known to the sender and receiver of data. The receiver transforms the received data using the decryption process to obtain the original data.

Symmetric key cryptography is a form of cryptosystem in which encryption and decryption are performed using the same key. It has been utilized for secure communications for long period of time.Symmetric key cryptography comprises two different methods for encryption and decryption. It can either use stream cipher or block cipher method of encryption/decryption.

Manuscript received Jun 12, 2011.

S.Anandi Reddy, Dhanalakshmi Srinivasan Engineering College, Perambalur, India. (e-mail : anandinearu@gmail.com)

1.2 EVOLUTION OF ADVANCED ENCRYPTION STANDARD

The National Institute of Standards and Technology initiated a process to select a symmetric key encryption/ decryption algorithm in 1997. Finally, Rijndael algorithm was accepted among other finalists as the Advanced Encryption Standard (AES) in 2001.

It is noted that before the acceptance of Rijndael algorithm, DES and its improved variant 3DES were used as symmetric key standards. DES has 16 rounds and encrypts and decrypts data using a 64-bit key. 3DES has hardware implementation that doesn't produce efficient software code and three times as many rounds as DES so correspondingly slower. Both DES and 3DES use a 64-bit block size. To satisfy both efficiency and security, a larger block size is desirable.

AES-128 has 10 rounds where data is encrypted and decrypted in 128-bit blocks using a 128-bit key. It is a very good performer in both hardware and software across a wide range of computing environments.

1.3 MOTIVATION

The objective in using AES is to transfer the data so that only the desired receiver with a specific key would be able to retrieve the original data. However, the natural and malicious injected faults reduce its reliability and may cause confidential information leakage. Thiscan be either due to:

- Natural faults caused by defects in gates or,
- Malicious injected faults to retrieve the key and break the system.

As a result, finding a suitable fault detection scheme has always been an issue in the AES. FPGAs are most flexible implementation to produces high performance withlow cost. FPGA provides more physical security with parallelism.

1.4 OBJECTIVES

The objective of this paper is to find concurrent structure independent fault detection schemes for reaching reasonable fault coverage. It makes a robust implementation of AES against these above attacks and provides highest efficiencies, showing reasonable area and time complexity overheads.

II. ADVANCED ENCRYPTION STANDARD

2.1 AES ALGORITHM

AES is an iterated block cipher with a fixed block size of 128 and a variable key length. It hasvariable number of rounds, which is fixed according to key length. AES performs four transformations in each round in order to provide high level of security.

2.2 TRANSFORMATIONS

AES performs four transformations in each round in order to provide high level of security. This involves the properties of confusion and diffusion to provide frustrating statistical cryptanalysis. The transformations in each round of encryption except for the last round are as follows:

I. **SubBytes:** It isa non-linear substitution step where each byte is replaced with another according to a lookup table. The look table is known as S-Box which is generated by applying affine transform to multiplicative inverse of input.

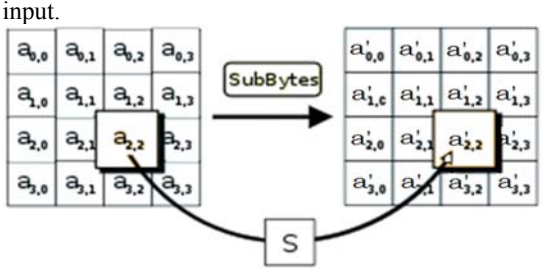


Figure.1

II. **ShiftRow:** It is a transposition step where each row of the state is shifted cyclically a certain number of steps to the left.For AES, the first row is left unchanged. Each byte of the second row is shifted one to the left. Similarly, the third and fourth rows are shifted by offsets of two and three respectively. Similarly for decryption rows are shifted right.

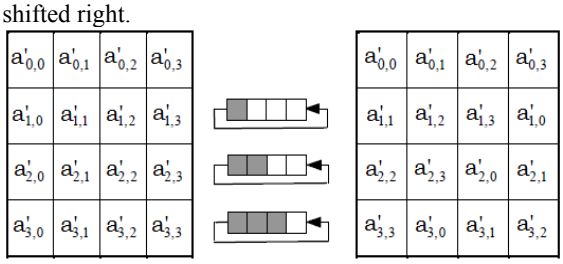


Figure.1.1

III. **MixColumn:** It is a mixing operation which operates on the columns of the state, combining the four bytes in each column using an invertible linear transformation

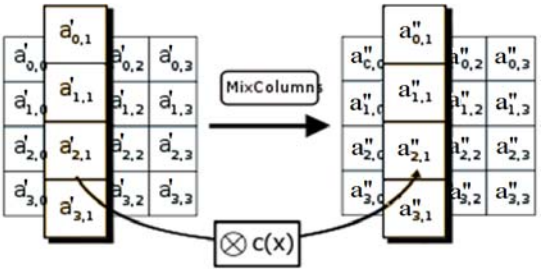


Figure.1.2

During this operation, each column is multiplied by the known matrix that for the 128 bit key is

$$\begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 3 \\ 3 & 1 & 1 & 2 \end{bmatrix}.$$

IV. **AddRoundKey:** In this, each byte of the state is combined with the round key; each round key is derived from the cipher key using a key schedule. The roundKey is added to the state before starting theloop. In the AddRoundKey step, each byte of the state is combined with a byte of the round sub key using the XOR operation.

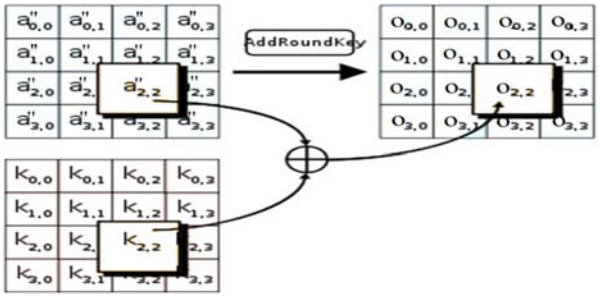


Figure.1.3

III. EXISTING SYSTEM

There has been many fault detection schemes proposed till this date to avoid the possibility of suffering from various attacks such as natural faults caused by defects in gates ,injection of fault by attackers to retrieve the key. Some of the majorly contributed schemes follow as:

3.1 A 16-BIT KEY PARITY METHOD

In this scheme, the output parity bits of each transformation in every round are predicted from the inputs. Then, the comparisons between the predicted and the actual parities (obtained using the actual parity block or predetermined parity block) can be scheduled so that the desired error coverage is obtained. Since the 128 bit input is represented in 4X4 matrix 16 parity bits corresponding to each 1 byte are compared which is presented in [3] and [12]. It has drawback that requires two blocks of 256 x 9 memory cells (S-boxes and parity predictions box). So it has relatively high area complexity for the parity predictions of MixColumns in the AES encryption. This is even more for Inv MixColumns in the AES decryption.

3.2 REDUNDANCY-BASED TECHNIQUE

The redundancy-based solution for implementing fault detection in the encryption module is based on the idea of performing a test decryption immediately after the encryption and then checking whether the original data block is obtained. The redundant unit fault detection scheme [4], [6] is used where algorithm-level, round-level, or operation-level fault detections are considered. The schemes pays time penalty either to decrypt a data block or for the comparison.

3.3 DOUBLE TIME TRANSFORMATION TECHNIQUE

In [5], the scheme uses same transformations twice in an AES round for the same data to detect the transient errors. It is time consuming and hence increases delay overhead. However, this method suffers from permanent internal faults or the malicious injected faults lasting for a long period.

3.4 MULTIPLICATION-BASED SCHEME

In [7] scheme, the result of the multiplication of the input and the output of the multiplicative inversion is compared with the predicted result of unity.

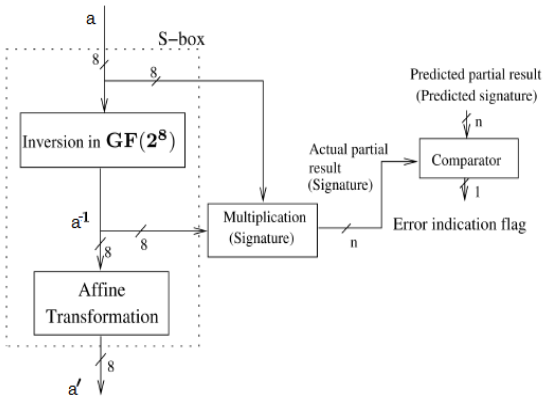


Figure. 2 The multiplication-based scheme for the fault detection of the multiplicative inversion

Since S-box is generated by applying affine transform to multiplicative inverse of an input, there is no access to the output of the multiplicative inversion. Therefore, this scheme is not suitable for the S-boxes and inverse S-boxes implemented using lookup tables (LUTs).

3.5 Pipelined Structure

Under this, pipelined distributed memories for the LUT-based S-boxes and inverse S-boxes are used to increase the design speed and the overall frequency of AES. There are three architecture to speed up, namely pipelining, subpipelining, and loop unrolling. Among these approaches, the subpipelined architecture can achieve maximum speed up and optimum speed–area ratio in non-feedback modes. Subpipelining inserts rows of registers among combinational logic not only between but also inside each round unit which is presented in [8].Non-LUT-based approaches can be used to avoid the unbreakable delay of LUTs which involve inversions in Galois Field, which may have high hardware complexities.

IV. PROPOSED SYSTEM

The proposed system detects fault while implementing a cipher from a plaintext for transmission and thus called concurrent error detection (CED).

4.1AES Encryption

The new fault detection structure for the AES encryption consist of two sets of error indication flags corresponding tocombined SubBytes and ShiftRows and combined MixColumns and AddRoundKey. A typical AES encryption round (except for the last round) consists of four transformations, and the fault detection schemes are follows.

4.1.1 SubBytes and ShiftRows

In the AES encryption, the SubBytes transformation consists of 16 S-boxes corresponding to 16 one byte of 128 bit input. Let $e_{r,c}$, $0 \le r,c \le 3$, be the error indication flag for the S-box with the input $a_{r,c}$ and the output $a'_{r,c}$.The output state of such flags can be written as 16 formulations as follows:

$$e_{r,c} = P_{(M_{r,c}a'_{r,c}+m_{r,c})} + u'_{r,c}; \quad 0 \le r,c \le 3,$$

(5)

where $u'_{r,c} = (u', 0,0,0,0,0,0,0)^T$, u is obtained by logical OR operations of all inputs and outputs of S-Box, i.e.,

$$u' = (a_0 \cup a_1 \cup \ldots\ldots a_7) \cup (\overline{a'_0} \cup \overline{a'_1} \cup a'_2 \cup a'_3 \cup a'_4 \cup \overline{a'_5} \cup \overline{a'_6} \cup a'_7)$$

. For input hex (a) =00 and output hex (a') =63 of S-Box, u'=0 but for remaining input u'=1.And we have

$$Ma' + m = u' \qquad (6)$$

where M is 8 X 8 matrix

$$M = \begin{bmatrix} a_{6,5,2} & a_{5,4,1} & a_{7,5,3,0} & a_{6,4,2} & a_{7,5,3,1} & a_{7,6,5,2,0} & a_{7,6,5,4,1} & a_{7,6,3,0} \\ a_{7,5,3,2,0} & a_{6,4,2,1} & a_{7,6,5,4,3,1} & a_{7,6,5,4,3,2,0} & a_{7,6,5,4,3,2,1} & a_{5,3,2,1} & a_{4,2,1,0} & a_{6,4,3,1} \\ a_{6,4,3,1} & a_{7,5,3,2,0} & a_{7,6,5,4,3,2} & a_{7,6,5,4,3,1} & a_{7,6,5,4,3,2,0} & a_{6,4,3,2} & a_{5,3,2,1} & a_{7,5,4,2,0} \\ a_{7,6,4,0} & a_{6,5,3} & a_{6,0} & a_{7,5} & a_{6,4} & a_{6,4,3,2,0} & a_{7,5,3,2,1} & a_{7,5,1} \\ a_{7,6,2,1} & a_{7,6,5,1,0} & a_{5,3,1} & a_{4,2,0} & a_{3,1} & a_{6,4,3,2,1} & a_{7,5,3,2,1,0} & a_{7,3,2} \\ a_{7,3,2} & a_{7,6,2,1} & a_{6,4,2,0} & a_{5,3,1} & a_{4,2,0} & a_{7,5,4,3,2} & a_{6,4,3,2,1} & a_{4,3,0} \\ a_{4,3,0} & a_{7,3,2} & a_{7,5,3,1} & a_{6,4,3,0} & a_{5,3,1} & a_{6,5,4,3,0} & a_{7,5,4,3,2} & a_{5,4,1} \\ a_{5,4,1} & a_{4,3,0} & a_{6,4,2} & a_{7,5,3,1} & a_{6,4,2,0} & a_{7,6,5,4,1} & a_{6,5,4,3,0} & a_{6,5,2} \end{bmatrix}$$

and

$$m = \left[a_{6,0}, a_{7,6,1}, a_{7,2,0}, a_{6,3,1}, a_{7,6,4,2}, a_{7,5,3}, a_{6,4}, a_{7,5}\right]^T$$

Therefore, $P_{(M_{r,c}a'_{r,c}+m_{r,c})}$ is presented as

$$P_{(Ma'+m)} = a_0(a'_x + a'_y) + a_1 a'_x + a_2 a'_z + a_3 a'_4 + a_4(a'_y + a'_3) + a_5 a'_w + a_6(a'_z + \overline{a'_6}) + a_7(\overline{a'_5} + a'_4) = u'$$

(7)where

$$a'_w = a'_0 + a'_2 + a'_3 + a'_5, a'_x = a'_w + a'_7, a'_y = a'_1 + a'_4 + a'_5 \text{ and } a'_z = a'_2 + a'_7.$$

The 128-bit output of the SubBytes transformation acts as the input to ShiftRow, the output state of ShiftRows is obtained by just shifting the state entries in its input state. Hence the state entries in each row remain the same but differ by location.Therefore, by considering the output of ShiftRows and equation (5), for row r and column c, the output state of the flags can be rewritten as 16 formulations as follows:

$$e_{r,c} = P_{(M_{r,c^*}a'_{r,c^*}+m_{r,c^*})} + u'_{r,c^*}; \quad 0 \le r,c \le 3,$$

(8)

where c*= (r + c) mod 4. There by 16 error indication flags is generated from the output of ShiftRows for two transformations of SubBytes and ShiftRows together, i.e., one error indication flag for each byte, are obtained. This is shown in Fig.4.

4.1.2 MixColumns and AddRoundKey

The next two transformations in a typical AES encryption round are MixColumns and AddRoundKey. The MixColumns and AddRoundKey transformations are constructed matrix multiplication and modulo-2 addition of the input state with the roundkey respectively. Here low-complexity fault detection scheme derived for combined transformation of MixColumns and AddRoundKey.

Let $SR(A') = \left[a'_{r,c^*}\right]_{r,c=0}^{3}$ and $K - \left[k_{r,c}\right]_{r,c=0}^{3}$ be the input and the roundkey input of MixColumns and AddRound- Key in round i, respectively. Let the output of AddRoundKey be $O = \left[o_{r,c}\right]_{r,c=0}^{3}$ .Then, the following holds:

$$\sum_{r=0}^{3} o_{r,c} = \sum_{r=0}^{3} a'_{r,c*} + \sum_{r=0}^{3} k_{r,c}, \quad 0 \le c \le 3. \quad (9)$$

where c*= (r + c) mod . This can be rewritten as

$$\sum_{r=0}^{3}\left(a'_{r,c*} + k_{r,c} + o_{r,c}\right) = 0 \in GF(2^8), \quad 0 \le c \le 3 \quad (10)$$

and each summation is over GF (2^8) which consists of eight modulo-2 additions.

$$\sum_{r=0}^{3} a''_{r,c} = \sum_{r=0}^{3} a'_{r,c*}, \quad 0 \le c \le 3. \quad (11)$$

And we have

$$\sum_{r=0}^{3} o_{r,c} = \sum_{r=0}^{3} a''_{r,c} + \sum_{r=0}^{3} k_{r,c}, \quad 0 \le c \le 3 \quad (12)$$

Therefore,

$$\sum_{r=0}^{3} o_{r,c} = \sum_{r=0}^{3} a'_{r,c*} + \sum_{r=0}^{3} k_{r,c}, \quad 0 \le c \le 3 \quad (13)$$

Now let us introduce the four 8-bit error indication flags for four columns of the state as

$$E_c = \sum_{r=0}^{3}(a'_{r,c*} + k_{r,c} + o_{r,c}), \ 0 \le c \le 3 \quad (14)$$

In error-free situation, by using (10) all 32 bits of such flags in (14) must be zero, i.e. $E = 0 =$(0, 0… 0) $\in$GF (2^3), $0 \le c \le 3$.

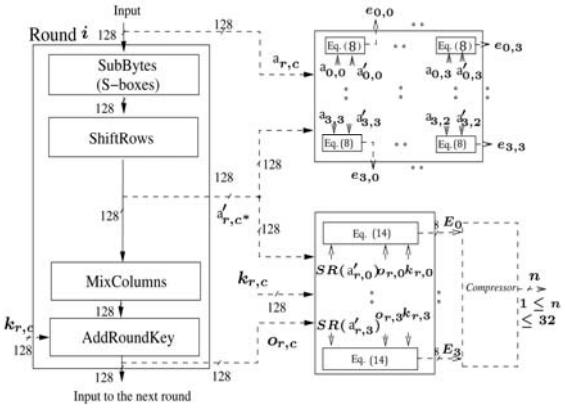


Figure.3 The proposed fault detection scheme for the i[th] round of the AES encryption.

These 32 error indication flags can be used for these two combined transformations, i.e., eight error indication flags for each column of the state matrix as shown in Fig. 4.1. This error indication flags can be compressed so that n, $1 \le n \le 32$, error indication flags for these two transformations are achieved. This can be performed by ORing different combinations of the 32 error indication flags obtained in (14) as denoted by the compressor block in Fig.4. With 16 error indication flags by compression, greater than 99 percent of the errors are covered.

The last round of the every AES encryption (10[th] round in AES-128 encryption) consists of all transformations, (SubBytes, ShiftRows, and AddRoundKey) except MixColumns transformation. Similar to all other rounds of the AES encryption, 16 error indication flags for SubBytes and ShiftRows combined can be used for the last encryption round. Consequently, (14) can also be used for the last round.

4.1.3 Further Enhancement

The complexity of the scheme is reduced by modifying the structure using subexpression sharing. This reduces the number of logic gates utilized in obtaining two sets of the error indication flags to have low-complexity fault detection scheme of the AES encryption, as shown in Fig.5. This is performed by modulo-2 addition of two sets of four coordinates of (14) for each column, i.e., $E_c = (e_{c,7}, e_{c,6}, \ldots, e_{c,0}) \in GF(2^8)$, $0 \le c \le 3$. Let $\hat{E}_c = (e_{c,4}, e_{c,2}, e_{c,1}, e_{c,0})$ and $\check{E}_c = (e_{c,5}, e_{c,7}, e_{c,6}, e_{c,3})$. Then, the four error indication flags of column c of the state are $\bar{E}_c = \hat{E}_c + \check{E}_c$; $0 \le c \le 3$. (15)

One can utilize four sets of modulo-2 additions of the output bits of each S-box pre-computed in (7), i.e., $a'_4 + a'_5, a'_2 + a'_7, a'_1 + a'_6$ and $a'_0 + a'_3$, to obtain the low-complexity error indication flags in (14). We use (14) to derive 16 low complexity signatures for the MixColumns and AddRound-Key transformations, i.e., four signatures for each column of the state matrix. This is shown in Fig. 5.This proposed fault detection for the MixColumns transformation which has 25 percent less areas overhead than the parity based scheme.

In fig 5, the Common Subexpressions (CSs) unit has been utilized to obtain 64 common subexpressions, i.e., 4 for each of the 16 S-boxes in the SubBytes transformation. If any of the two derived sets of error indication flags are one, the error is detected whereas if all of them are zero then no error has been detected although the output can be erroneous or correct. The (8) utilizes the hardware implementation of (7) which isless

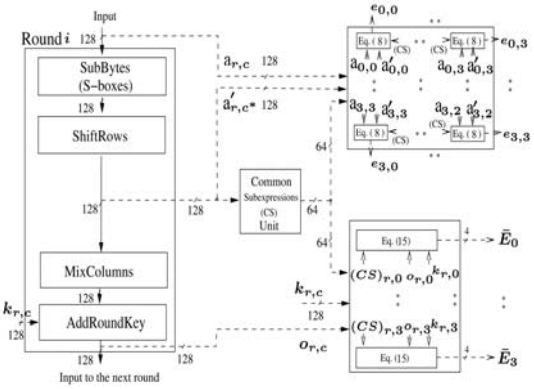


Figure.4 The low-complexity fault detection scheme utilizing subexpression sharing.

complex when the common sub expressions are used. Hence the Fig.5 shows less complexity compared to Fig.4.

4.2 AES Decryption

The AES decryption rounds also (except for the last round) consist of four transformations, i.e., InvShiftRows, InvSubBytes, AddRoundKey, and InvMixColumns. All the steps is similar to encryption but in reverse manner.

4.2.1 Inv Shift Rows and Inv Sub Bytes

In the AES decryption, the 128-bit input to InvShiftRows, i.e., the state matrix S' entries, is cyclically shifted to the right with the first row remaining unchanged.

$$e_{r,c} = P_{(M_{r,c}a'_{r,c^*}+m_{r,c})} + u'_{r,c}; \quad 0 \le r, c \le 3,$$

where c*= |r – c|. (16)

According to (15), 16 error indication flags for the Inv Shift Rows and Inv Sub Bytes transformations are generated.

4.2.2 ADDROUNDKEY AND INVMIXCOLUMNS

In decryption, InvMixColumns transformation is equivalent to multiplying the input state with the constant output matrix. In the AddRoundKey transformation, the input state, i.e., S, is added with the roundkey input state,i.e., K.

$$E_c = \sum_{r=0}^{3}\left(a_{r,c} + k_{r,c} + o_{r,c}\right), \quad o \le c \le 3$$

(17)

As in case of encryption, decryption also has three rounds in its last round i.e. InvMixColumns is removed. Similarly same (15) and (16) can be used in last round to detect fault.

For decryption also by using subexpression sharing the area overhead have been reduced 64 XOR gates to 48 XOR i.e. reduced by25 percent. Then, the four error indication flag for column c of state are

$$\bar{E}_c = \hat{E}_c + \hat{E}_c \;, 0 \le c \le 3 \qquad (18)$$

where $\hat{E}_c$ = ($e_{c,3}$, $e_{c,2}$, $e_{c,1}$, $e_{c,0}$) and $\hat{E}_c$= ($e_{c,7}$, $e_{c,6}$, $e_{c,5}$, $e_{c,4}$). The proposed scheme has less area and critical path delay when compared to other schemes presented for InvMixColumns. It requires 48 XOR gates with two XOR in critical path delay. Overall 25 percent area overhead and 33 percent in critical path delay has been reduced in proposed scheme.

4.3 ERROR SIMULATION

When exactly 1 bit error appears at the output of the AES encryption or decryption rounds, the parity-based fault detection scheme is able to detect it and the error coverage will be 100 percent. But when there is case of multiple errors, the results of our simulations are valid.

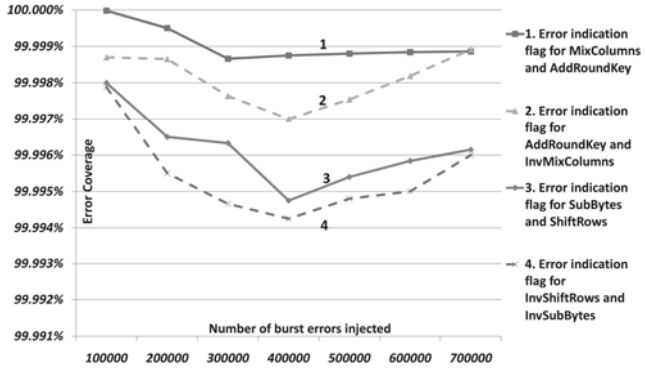


Figure.5 Simulation result for error coverage

We have considered both single and multiple stuck-at errors for the proposed scheme. These models cover both natural faults and fault attacks. In our simulations, we injected errors in two manners, i.e., burst and random errors, and obtain the error coverage for these two cases:

Burst Error:

In this type, the errors are injected at the 128-bit output of only one transformation in the AES encryption /decryption. The errors are monitored by injecting burst errors one at a time up to 700,000 at the transformation outputs. The error coverage for the two sets of error indication flags is greater than 99.996 percent.

Random errors:

This type of errors is injected at random locations, i.e., four 128-bit outputs of the transformations. The errors are covered either by one of the two series of the error indication flags.

The increase in the number of error injected increases the error coverage close to 100 percent. In Fig.7, the solid and dashed lines represent the error coverage for the AES encryption and decryption, respectively. For certain AES implementations containing storage elements, one can use the error correcting code-based approach presented in [13] in addition to the proposed scheme in this paper to make a more reliable AES implementation.

V. CONCLUSION

In this paper, we have considered a structure independent fault detection scheme for the AES encryption and decryption. This can be applied for both the S-boxes and the inverse S-boxes using lookup tables and those utilizing logic gates based on composite fields.using S-boxes and inverse S-boxes used for both LUT and composite fields. The proposed scheme has been simulated and its fault coverage has been evaluated in detail. The proposed system is able to find the round and its corresponding transformation in which fault occurred. Thereby optimized hardware is achieved bymodifying the structure using subexpresssion sharing. Hence the reduced number of gates is required in the implementation of AES.The slice overheads are less than those for the other schemes which have the same error coverage. Thus, this scheme has the highest efficiencies, showing reasonable area and time complexity overheads. Hence the proposed schemes outperform the previously reported ones.

VI. ACKOWNLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to all those who made the achievement possible without whom it would never have come into existence. I express my gratitude and thanks to my parents first for giving health as well as sound mind &financial support for taking up this paper. I would like to express my sincere appreciation and gratitude to Assit.Prof. M.Arul Kumar for his supervision and guidance during my studies and to develop this paper.

REFERENCES

[1] M. Mozaffari-Kermani and A. Reyhani-Masoleh, Member, “Concurrent Structure-Independent Fault Detection Schemes for the Advanced Encryption Standard,” IEEE Transactions on Computers, Vol. 59, No. 5, May 2010.

[2] C. Moratelli, F. Ghellar, E. Cota, and M. Lubaszewski, “A Fault-Tolerant DFA-Resistant AES Core,” Proc. IEEE Int’l Symp. Circuits and Systems (ISCAS ’08), pp. 244-247, May 2008.

[3] R. Karri, G. Kuznetsov, and M. Goessel, “Parity-Based Concurrent Error Detection of Substitution-Permutation Network BlockCiphers,” Proc. Int’l Workshop Cryptographic Hardware and Embedded Systems (CHES ’03), pp. 113-124, Sept. 2003.

[4] R. Karri, K. Wu, P. Mishra, and K. Yongkook, "Fault-Based Side Channel Cryptanalysis Tolerant Rijndael Symmetric Block Cipher Architecture," Proc. IEEE Int'l Symp. Defect and Fault Tolerance in VLSI Systems (DFT '01), pp. 418-426, Oct. 2001.

[5] T.G. Malkin, F.X. Standaert, and M. Yung, "A Comparative Cost/Security Analysis of Fault Attack Countermeasures," Proc. Int'l Workshop Fault Diagnosis and Tolerance in Cryptography (FDTC '06),pp. 159-172, Oct. 2006.

[6] C.H. Yen and B.F. Wu, "Simple Error Detection Methods for Hardware Implementation of Advanced Encryption Standard," IEEE Trans. Computers, vol. 55, no. 6, pp. 720-731, June 2006.

[7] M. Karpovsky, K.J. Kulikowski, and A. Taubin, "Differential Fault Analysis Attack Resistant Architectures for the Advanced Encryption Standard," Proc. Conf. Smart Card Research and Advanced Applications (CARDIS '04), vol. 153, pp. 177-192, Aug. 2004.

[8] X. Zhang and K.K. Parhi, "High-Speed VLSI Architectures for the AES Algorithm," IEEE Trans. Very Large Scale Integration Systems, vol. 12, no. 9, pp. 957-967, Sept. 2004.

[9] P. Maistri and R. Leveugle, "Double-Data-Rate Computation as a Countermeasure against Fault Analysis," IEEE Trans. Computers, vol. 57, no. 11, pp. 1528-1539, Nov. 2

[10] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "A Structure-Independent Approach for Fault Detection Hardware Implementations of the Advanced Encryption Standard," Proc. Int'l Workshop Fault Diagnosis and Tolerance in Cryptography (FDTC '07), pp. 47-53, Sept. 2007.

[11] C. Moratelli, E. Cota, and M. Lubaszewski, "A Cryptography Core Tolerant to DFA Fault Attacks," Proc. Ann. Symp. Integrated Circuits and Systems Design (SBCCI '06), pp. 190-195, Sept. 2006.

[12] G. Bertoni, L. Breveglieri, I. Koren, P. Maistri, and V. Piuri, "AParity Code Based Fault Detection for an Implementation of the Advanced Encryption Standard," Proc. IEEE Int'l Symp. Defect and Fault Tolerance in VLSI Systems (DFT '02), pp. 51-59, Nov. 2002.

[13] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "A Lightweight Concurrent Fault Detection Scheme for the AES S-Boxes Using Normal Basis," Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES '08), pp. 113-129, Aug. 2008.

BIOGRAPHY

S Anandi Reddy received the BE degree in Electronics and Communication Engineering in 2008, and ME degree in Communication System in 2011 from Dhanalakshmi Srinivasan Engineering College. Hermain interests include computer architecture and VLSI chip design.

Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

Multidimensional Context Dependent Information Delivery on the Web

[1] V. Mani Sarma, [2] Prof.P.Premchand

Abstract— **Multidimensional Semi structured Data MSSD are semi structured data that Present deferent facets under deferent contexts i.e. alternative worlds For the representation of MSSD various formalisms have been proposed by the authors both syntactic such as MSSD expressions and MXML as well as graphical such as Multidimensional OEM In this paper we present an infrastructure for handling MSSD This infrastructure provides appropriate tools for building MSSD applications and is independent from any particular application that uses it We also present a graphical interface called MSSDesigner that provides access to the infrastructure and we describe OEM History an MSSD application that supports keeping track of temporal changes in semi structured databases**

Index Terms— **SSD, OEM Graph, OEM History, MDSS Data, MOEM Graph..**

I. INTRODUCTION

The nature of the Web poses a number of new problems While in traditional databases and information systems the number of users is more or less known and their background is to a great extent homogeneous Web users do not share the same background and do not apply the same conventions when interpreting data Such users can have deferent perspectives of the same entities a situation that should be taken into account by Web data models Those problems call for a way to represent information entities that manifest deferent facets whose contents can vary in structure and value Multidimensional Semistructured Data MSSD paired with an extension of OEM called multidimensional OEM have been proposed in MSSD and MOEM incorporate ideas from multidimensional programming languages and associate data with dimensions in order to tackle the aforementioned problems In MSSD variants of the same information entities each holding under a specific world have been consolidated to form multidimensional entities Syntactic expressions called context specifiers are associated to pieces of data facets of multidimensional entities and specify sets of worlds under which these data hold In this paper we present the overall architecture of an infrastructure that allows the management of multidimensional semistructured data This infrastructure can be used for the development of new applications and MSSD tools that will be placed on top of it by providing access to a number of operations on MSSD We focus mainly on MOEM graphs that can be used to represent MSSD and we present MSSDesigner a graphical interface for handling MOEM graphs which is also a part of the infrastructure A world represents an environment under which data obtain a substance In the following definition we specify the notion of world using a set of parameters called dimensions.

Definition: 1 Let D be a set of dimension names and for each d € D let Vd be the domain of d, with Vd ≠0. A world w with respect to D is a set whose elements are pairs(d v), where d € D and v € Vd such that for every dimension name in D there is exactly one element in w.

The main difference between conventional and multidimensional semi structured data is the in introduction of context specifieres. Context specifieres are syntactic constructs, expressing constraints on dimension values that are used to qualify semi structured data expressions (SSD-expressions) and specify sets of worlds under which the corresponding SSD-expressions hold. In this way it is possible to have at the same time variants of the same information entity, each holding under a different set of worlds. An information entity that encompasses a number of variants is called multidimensional entity, and its variants are called facets of the entity. The facets of a multidimensional entity may differ in value and or structure, and can in turn be multidimensional entities or conventional information entities. Each facet is associated with a context that defines the conditions under which the facet becomes a holding facet of the multidimensional entity. If a facet **f** of a multidimensional entity **e** holds under a world W (Or under every world defined by a context specifier c) then we say that **e** evaluates to **f** under **w** (under c, respectively).

Example1. The use of dimensions for representing worlds is shown with the following three context specifiers:

1. [time in {07:00..15:00}]
2. [language= English, detail in {low, medium}]
3. [Season in {fall, spring}, daytime= noon| season= summer]

In Example 1, context specifies (a) represents the worlds for which the dimension time can take any value between 07:00 and 15:00, while (b) represents the worlds for which language is English and detail is either low or medium. Context specifier (c) is more complex, and represents the worlds where season is either fall or spring and daytime is noon, together with the worlds where season is summer. Notice that according to Definition 1, for a set of (dimension, value) pairs to represent a world with respect to a set of dimensions **D** it must contain exactly one pair for each dimension in D Therefore if D f language detail g with V language f English g and V detail flow medium high g then

Manuscript received March 22, 2011.

V. Mani Sarma, Associate Professor, Holy Mary Institute of Technology & Science, Hyderabad, Andhra Pradesh, India-501 301,
(e-mail : Manisharma.vittapu@gmail.com)
Prof.P.Premchand, Department of CSE, Osmania University, Hyderabad, Andhra Pradesh, India. (e-mail : Drpremchand_p@yahoo.com)

f language English detail low g is one of the six possible worlds with respect to D This world is represented by context specifier b in Example together with the world f language English detail medium g . Notice that it is not necessary for a context specifier to contain values for every dimension in D Omitting a dimension implies that its value may range over the whole dimension domain. The context specifier is called universal context and represents the set of all possible worlds with respect to a set of dimensions D.

2.2 MULTIDIMENSIONAL OEM

Multidimensional Object Exchange Model (MOEM) is an extension of Object Exchange Model OEM suitable for representing multidimensional data. MOEM extends OEM with two new basic elements:

Multidimensional nodes represent multidimensional entities and are used to group together nodes that constitute facets p of such entities. Graphically multidimensional nodes have a rectangular shape to distinguish them from conventional circular nodes, which are called context nodes.

Context edges are directed labeled edges that connect multidimensional nodes to their facets. The label of a context edge pointing to a facet p is a context specifier defining the set of worlds under which p holds Context edges are drawn as thick lines to distinguish them from Conventional thin lined edges called entity edges. The definition of multidimensional data graph is given below

Definition :2 Let C be a set of context specifiers, L be a set of labels and A be a set of atomic values. A multidimensional data graph is a finite directed edge-labeled multigraph G=(V, E, r, C,L,A, v) where:

1. The set of nodes V is partitioned into multidimensional nodes and context nodes V= Vmld U Vcxt. Context nodes are further divided into complex nodes and atomic nodes Vcxt = Vc U Va.

2. The set of edges E is partitioned into context edges and entity edges E= Ecxt U Eett, such that Ecxt Vmld C V and Eett Vc LV

3. r € V is the root, with the property that there exists a path from r to every other node in V.

4. v is a function that assigns values to nodes, such that v(x) = M if x € Vmld, v(x)= C if x € Vc and v(x)= v^I (x) if x€ Va, where M and C are reserved values and v^I is a value function v^I: Va → A which assigns values to atomic nodes

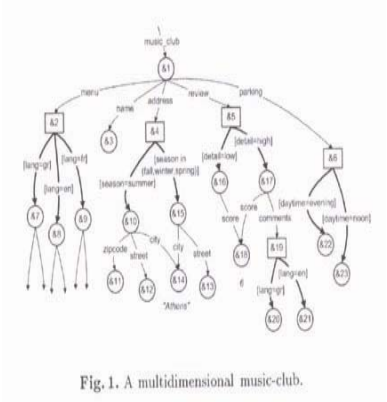


Fig. 1. A multidimensional music-club.

As an example consider the part of an MOEM graph in Figure which represents context dependent information about a music club. The graph is not fully developed and some of the atomic objects do not have values attached. The music club with oid &1 operates on a different address during the summer than the rest of the year in (Delhi it is not unusual for clubs to move south close to the sea in the summer period and north towards the city center during the rest of the year). Except from having a different value context objects can have a different structure as is the case of and which variants of the multidimensional object address with oid &4. The menu of the club is available in three languages namely English, French and Germen. In addition the club has a couple of alternative parking places, depending on the time of day as expressed by the dimension daytime. Two fundamental concepts related to multidimensional data graphs are the notions of explicit and inherited contexts. The explicit context of a context edge is the context specifier assigned to that edge, while the explicit context of an entity edge is the universal context specifier. The explicit context can be considered as the "true" context only within the boundaries of a single multidimensional entity .When entities are connected together in an MOEM graph, the explicit context of an edge is not the "true" context in the sense that it does not alone determine the worlds under which the destination node holds. The reason for this is that when an entity $\mathbf{e_2}$ is part of (pointed by through an edge) another entity $\mathbf{e_1}$ then $\mathbf{e_2}$ can have substance only under the worlds that $\mathbf{e_1}$ has substance. This can be conceived as if the context under which $_{e1}$ holds is inherited to $\mathbf{e_2}$. The notion of validity of an MOEM graph ensures that edges pointing to multidimensional nodes do not exist in vain in particular, an edge h leading to a node q is invalid if the inherited context of h has no common world with the context union of the worlds represented by the explicit contexts of the edges that depart from q.

2.3 MULTIDIMENSIONAL XML

Besides MOEM which models MSSD as a graph a notation for expressing MSSD has been also proposed. The notation extends ssd-expression with context specifiers and is called mssd-expression. Another way to describe MSSD is Multidimensional XML (MXML) which is an extension of XML that incorporates context specifiers. In MXML elements and attributes may depend on a number of dimensions. A multidimensional element is denoted by preceding its name with the special symbol "**©**" and encloses one or more context elements that constitute facets of that multidimensional element, holding under the worlds specified by the corresponding context specifier. Context elements have the same form as conventional XML elements. MXML suggests a new way for designing Web pages which encode context dependent data. The multidimensional paradigm allows a single document to have a number of variants each holding under a specific world. Information in such a document is encoded in MXML. An MXML document may be associated with a Multidimensional XSL style sheet MXSL in short containing instructions on how to present information in XML documents An MXSL style sheet encodes a set of conventional XSL style sheets each being the facet of the MXSL under a specific world For each possible world the holding XSL is applied to the holding XML to give the view of the information under that world.

III. ARCHITECTURE OF AN MSSD INFRASTRUCTURE

An MSSD infrastructure is a set of tools and processes that create manipulate and query MSSD and are used directly or by applications that need the support of an MSSD framework. This section presents such an infrastructure for manipulating multidimensional semistructured data which can also be used for implementing additional tools and applications. The infrastructure consists of the following components described in Figure2.

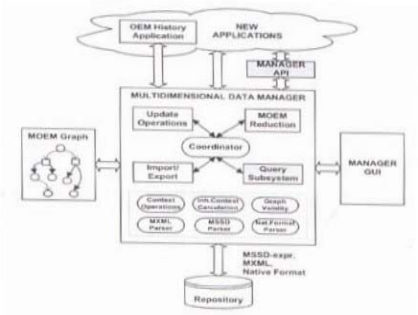


Fig. 2. Architecture of an MSSD infrastructure

MOEM Graph consists of the main memory data structures which actually hold graph representations of MSSD.

Multidimensional Data Manager (MDM) is responsible for managing MOEM graphs. It comprises a set of modules that allow the creation maintenance and querying of multidimensional semistructured data various modules of MDM can be accessed through graphical user interfaces offered by the Manager GUI.

Manager GUI comprises a number of user interfaces, which provide access to various functions of MDM, like MOEM graph creation and maintenance, and MOEM graph querying. MOEM graph creation and maintenance can be performed through MSSDesigner.

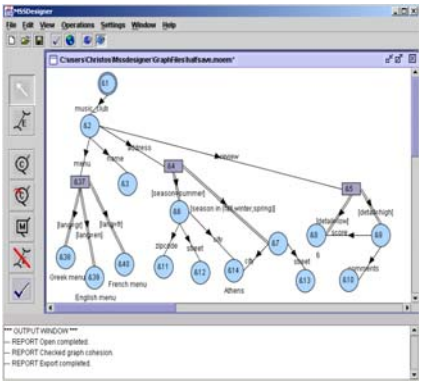


Figure.3.A Sample Image of MSSDesigner

Repository is the physical storage medium that supports the MDM needs for loading and saving MSSD and MOEM graph representations. Note that a number of formats able to represent semistructured data can be used when storing MOEM in files. At this moment, mssd-expressions, MXML and native format expressions are supported. **Manager API** aims at providing an application programming interface for new applications that will need to use the functionality of the system. This module enables applications to use the existing infrastructure by issuing commands in an especially made script like language. However an application can directly use the MDM as is the case of OEM History.

Multidimensional Data Manager MDM is the most important component It comprises a set of utility processes which appear inside a box placed at the bottom of MDM in Figure and are accessible to all other MDM modules Those utility processes are explained below

MSSD-Expressions The grammar of mssd-expressions is given in Extended Backus-Naur Form EBNF. Here we give as an example the mssd- expression that describes the address object with oid **&4** in Figure 1.

&4 ([seasonsummer]:
&10{zipcode:&11,street:&2, city:&14 " Delhi"}, [season in {fall, winter, spring}]: &15 { city: &14,street: &13})

MXML Representations MXML has been defined in the following MXML extract describes the same address object as the above mssd- expressions example:

```
<@address>
[season= summer]
        <address>
        <zipcode>…</ zipcode>
        <street>…</ street>
        <city id="c1"> Delhi <city>
</address>

[season in {fall, winter, spring}]
<address>
<city idref="c1" /> Delhi <city>
<street>…</ street>
</address>
[/]
</@address>
```

4.4 MOEM REDUCTION MODULE

This module is responsible for two jobs:

(a) Reduction of a MOEM graph to a conventional OEM graph holding under a specific world, and

(b) Partial reduction of a MOEM graph to another MOEM graph holding under a set of worlds. Reduction to OEM Given a specific world it is always possible to reduce an MOEM graph to a conventional OEM graph holding under that world. By specifying different worlds, the same MOEM can be reduced to different OEMs. The graph to be reduced must be context deterministic i.e for every multidimensional entity in the graph the context specifiers of that entity must be mutually exclusive. This ensures that two different facets of a multidimensional entity cannot hold under the same world. A procedure which performs reduction to OEM is presented below, and it is based on the idea that inherited contexts identify the parts of the graph that do not hold under a world.

The facet of an MOEM graph G under a world w, is an OEM graph G_w that holds under w Given a world w expressed as a context specifier c_w, the graph G_w can be obtained from G through the following process:

Procedure reduce to OEM (G, c_w G_w) is G_w

Step1: Remove every node and edge with $cw \cap i_c = 0c$, where i_c gives the inherited context of the node or edge respectively.

Step2: For every entity edge (p, l, m_1) with m_1 a multidimensional node, follow the path of consecutive context edges (m_1, c1,m2)…, (m_n, cn, q), n ≥1, until no more context edges can be followed. Then if q is a context node add a new entity edge(p, l, q) in the set of entity edges.

Step3: Remove all multidimensional nodes. Remove all edges departing from or leading to the removed nodes.

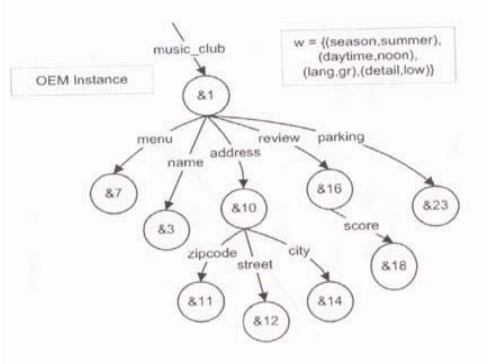


Figure.4 The OEM instance, holding under the world w, of the MOEM graph in Figure1.

Partial Reduction Partial reduction is in fact a generalization of the procedure reduce- to- OEM given above. In partial reduction a context specifier that represents a set of world's nearly more than one world is given. The MOEM graph is reduced to a new MOEM graph containing only the nodes and edges that hold under any of the specified worlds. In order to obtain the reduced MOEM graph the inherited context of nodes and edges is used, and a process similar to step1 of reduce- to- OEM is performed.

5.MSSDESIGNER

MSSDesigner is a graphical interface (part of the Manager GUI) that gives access to the functionality of MDM. A sample image of MSSDesigner Displaying a simple graph about a multidimensional music club is depicted in Figure4. MSSDesigner employs a multi document interface (MDI) where each document- frame corresponds to a data graph. All the operations performed by the various control buttons of the application, have effect to the currently focused frame. Through MSSDesigner it is possible to import a graph from an MSSD expression or MXML representation, and export a graph to one of those formats.

CONCLUSION

In this paper we proposed an architecture for manipulating MSSD that can be used as an infrastructure for the development of new MSSD tools and applications We showed the capabilities of this infrastructure and we presented MSSDesigner a graphical user interface for designing MOEM graphs that is a part of the GUI of this infrastructure Furthermore we explained how a new application can exploit this functionality and aims at accommodating temporal changes in semi structured databases. We believe that MOEM has a lot of potential and can be used in a variety of edges among which in information integration for modeling objects whose value or structure vary according to sources in digital libraries for representing metadata that conform to similar formats in representing geographical information where possible dimensions could be scale and theme.

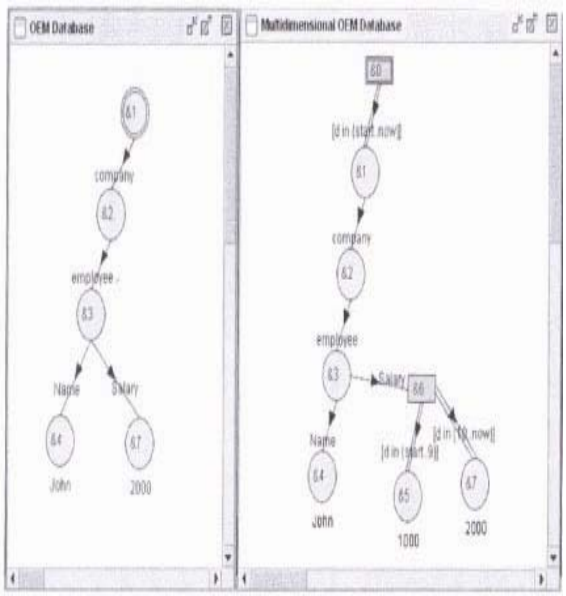


Figure.5 The Database after the insertion of new employee at d = 20

REFERENCES

[1] S Abiteboul P Buneman and D Suciu Data on the Web From Relations to Semistructured Data and XML Morgan Kaufmann Publishers2000.

[2] S Abiteboul D Quass J McHugh J Widom and J L Wiener The Lorel Query Language for Semistructured Data International Journal on Digital Libraries,1:68-88, 1997.

[3] Ph A Bernstein M L Brodie S Ceri D J DeWitt M J Franklin H GarciaMolina J Gray G Held J M Hellerstein H V Jagadish M Lesk D Maier J F Naughton H Pirahesh M Stone braker and J D Ullman The Asilomar Report on Database Research SIGMOD Record,27:74-80, 1998.

[4] S S Chawathe, S Abiteboul, and J.Widom, Managing historical semistructured data Theory and Practice of Object Systems24:1-20,1999.

[5] M Gergatsoulis, Y. Stavrakas, D. Karteris, A. Mouzaki and D. Sterpis, A Webbased System for Handling Multidimensional Information through MXML

[6] Accommodating changes in semistructured databases using multidimensional OEM in Advances in Databases and Information Systems.

[7] Information Systems Engineering TorontoOntario Canada May Lecture Notes in Computer Science LNCS Vol pages 183-199, springer-Verlag,2002.

BIOGRAPHY

I **Mani Sarma. V** received degree in Master of Computer Applications (MCA) from madras university in 1998, Chennai, Tamil Nadu, India and Master of Technology in Computer Science (M.Tech(CS)) from IETE in 2010, New Delhi respectively and pursuing Ph.D Degree in Computer Science since 2008 from Acharya Nagarjuna University

(ANU), Guntur, Andhra Pradesh, India. Since 1998, I have been working as a faculty member since 1999 in HITS Group and Associate professor since 2008 (College of Engineering) in Department of Computer Science and Engineering, Jawaharlal Nehru Technological University (JNTU), Andhra Pradesh, India. My Research interest includes Data Warehousing & Data Mining, Parallel and Distributed Mining, Distributed Data mining and Advanced Databases Systems. In my research experience I was published Two International journals and participated and presented Four National Conference papers were published from various engineering colleges in India. Two books (Study Material) were published by Tech publications and Spectrum Series for B.Tech (CSE) & IT III Year and II Year Students and also for MCA Students.

Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

A Study on Detection of Focal Cortical Dysplasia Using MRI Brain Images

[1] Dr.P.Subashini, [2] Ms.S.Jansi

***Abstract*— Focal Cortical Dysplasia (FCD) is the most frequent malformation of cortical development in patients with medically intractable epilepsy. In this paper, following brief introduction to the FCD, the chronology of its detection method is comprehensively surveyed. Next, the various techniques for detection of FCD are studied separately and their important factor and parameters are summarized in comparative table. It is the purpose of this paper to present an overview of previous and present conditions of the detection of FCD as well as its challenges. Accordingly, the importance, characteristics and the different approaches are discussed and analyses of these methods are evaluated.**

***Index Terms*— Focal Cortical Dysplasia (FCD), MRI, Gray-White Matter, Texture Analysis, Morphological operations.**

I. INTRODUCTION

Focal Cortical Dysplasia, a malformation caused by abnormalities of cortical development has been increasingly recognized as an important cause of medically intractable focal epilepsy. FCD was described as a pathologic entity first in 1971 by Taylor et al. FCD lesions are characterized on T1 weighted MRI by cortical thickening, blurring of GM/WM interface, and hyper intensity signal with respect to the rest of the cortex. Small FCD lesions are difficult to distinguish from non-lesional cortex and remain overlooked on radiological MRI inspection. Magnetic Resonance Imaging (MRI) plays a pivotal role in the presurgical evaluation of patients. MRI is currently the noninvasive method of choice for the in vivo diagnosis of FCD. Although MRI has allowed the detection of FCD in an increased number of patients, standard radiological evaluation fails to identify lesions in a large number of cases due to their small lesions and complexity of the cortex convolution [1].

Detecting the FCD, as epileptogenic lesion and consequently the decision about epilepsy surgery can never rely on one diagnostic tool alone. However, with respect only to brain imaging, MRI seems to be very important. In many patients, lesions of FCD are characterized by minor structural abnormalities that go unrecognized or are too subtle to be detected by standard radiological analysis. Using Quantitative methods, only few studies have been dedicated to the automatic detection of FCD and to the evaluation of structural changes too subtle can be detected by visual inspection. Niels K.Focke et al [2] presented a novel technique that uses standard clinical T_2 FLAIR scans to automatically detect FCDs. Leonardo Bonilha et al [3]; their work suggests that VBM (Voxel-Based Morphometry) can detect GMC excess in patients with FCD. The detection of FCD consists of several steps namely: preprocessing, enhancement, segmentation, feature extraction, and detection. After the detailed study of the previous research works on MRI brain images to detect the FCD, the various steps referred in the following figure Fig1, have to be proposed.

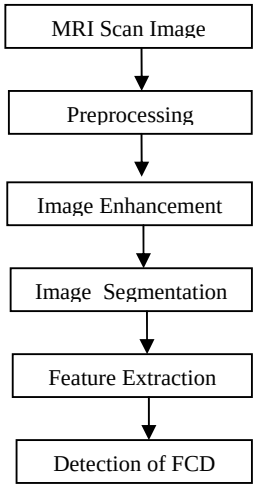


Figure.1 Proposed FCD detection

The rest of this paper is organized as follows: In section2, the overview of methodologies and technical details of previous work is described (i) Preprocessing: morphological operations are used; tissue classification is done. (ii) Image Enhancement: calculated the gray level intensity, smoothing and noise removal is done, and the threshold value is used to identify the lesion. (iii) Image Segmentation: segmenting the cortical tissues: WM, GM, and CSF. (iv) Feature extraction: calculating the color, texture, shape, and spatial relationship within the segmented model. DWT (Discrete Wavelet Transform) is used. (v) Detection of FCD: Automated classifier is used to identify FCD; MRI Characteristics of FCD are used to differentiate lesions from normal tissues. Finally, the conclusions are given in section3.

Manuscript received March 14, 2011.

Dr.P.Subashini, Associate Professor, Department of Computer Science and Engineering, Avinashlingam Deemed University for Women, Coimbatore, Tamilnadu, India – 641043. (e-mail : mail.p.subashini@gmail.com)

Ms.S.Jansi, Research Scholar, Department of Computer Science and Engineering, Avinashlingam Deemed University for Women, Coimbatore, Tamilnadu, India – 641043. (e-mail : jansi.sm@gmail.com)

II. METHODOLOGY

A. PREPROCESSING

The aim of preprocessing is to process the images in raw form and obtain images suitable for detection of FCD. All 3D MRI images are corrected for identifying non-uniformity, intensity standardization, automatic registration, automatic tissue classification, and Brain Extraction. Morphological operations such as dilation, erosion are used for removing the scalp and lipid layers. Cerebellum was also removed.

In 2002, Jan Kassubek, Hans-Jurgen Huppertz, et al.,[4] in their work based on using the SPM segmentation algorithm the gray matter was automatically segmented and the resulting gray matter was smoothed by using the fixed Gaussian kernel . Finally they represented the gray-matter density maps.

In 2005, Andy Khai Siang Eow [5] have proposed the different input modalities were considered for a particular patient and the tissue classification is done by considering the isotropic patient- specific head model.

In 2006, O. Colliot, T. Mansi et al [6] they used the Brain Extraction Tool (BET, Smith, 2002) for intensity non-uniformity and intensity standardization, automatic registration into a common stereotaxic space. For classifying the brain tissue in GM, WM and CSF the histogram method is used.

In 2009, Jeny Rajan, K.Kannan et al., [7] their work was based on the median voxel-wise intensity were normalized and morphological operations such as dilation, erosion and connected component analysis were used for removing the scalp and lipid layers from brain MR images. Reducing the false positives cerebellum was removed. The intensity threshold between gray and white matter was automatically determined by using the Gaussian curves. The white matter and CSF was removed from the segmented image.

In 2009, Rajeshwaran Logeswaran [8] has proposed to eliminate the background and artifacts by using the low-field MRI brain images in various regions. For identifying the WM, GM, ventricle, skull, etc., the Selection and Segmentation process were used and finally the MRI brain abnormalities were detected and labeled.

In 2009, April Khademi, Anastasios Venetsanopoulos, Alan Moody [9] have discussed to extract the entire brain region from FLAIR images various algorithms were required i.e. Global thresholding, Otsu thresholding, k-means clustering, active contours without edges and the BET tool were all unsuccessful. Firstly they applied a threshold value and then the absolute value was taken. Secondly, by applying a nonlinear mapping function they separated the intensity value (WML) from the outer head tissues. A k-means clustering is used to classify the regions and connected component analysis is used to find the largest region, which is the brain with WML included.

B. IMAGE ENHANCEMENT

The image enhancement is to improve the interoperability or perception of information in images for human viewers, or to provide 'better' input for other automated image processing techniques. Various noise and contrast with different percentages are generated. Noise level acquisition parameters have to be segmented. Threshold value is used to correctly identify the lesions.

In 2002, Jun Yang and Sung-Cheng Huang et al [10], work based on evaluation of different MRI segmentation approaches, the noise level of MR images varies with acquisition parameters including slice thickness, pixel size, field of view etc., An adaptive Gaussian noise distribution is assumed. Various noises with different percentages of signal power are generated using a Gaussian distribution random number generator and added to the simulated MR images.

In 2003, Andrea Bernasconi et al [11], their work proposed on advanced MRI for detection of FCD, to model the blurring of GM/WM transition, we calculated the absolute gradient of gray level intensities, a first-order texture feature. To model the hyper intense signal within the FCD on T1-weighted images, we developed and calculated the absolute difference between the intensity of a given voxel and the intensity at the boundary between GM and WM, defined using a histogram. To maximize the visibility of FCD lesions, a ratio map was generated.

In 2008, Pierre Besson, Olivier Colliot, Alan Evans et al [12], their work based on the automatic detection of FCD using surface-based features, the blurred WM/GM interface was modeled by applying a gradient operator on the MR image. The gradient magnitude was then interpolated at each vertex of the inner cortical surface to obtain the gradient surface map. The lesional probability maps obtained from the classifier were binarized by thresholding them at the best trade-off between detection rate and amount of false positives (FP). Using this threshold, the classifier correctly identified the lesion in 17/19 (89%) patients.

In 2008, Shan Shen, Andre J. Szameitat, and Annette Sterr [13], their work proposed on detection of infarct lesions from single MRI modality, the fuzzy memberships for each cluster are smoothed with a Gaussian kernel of 4mm to increase connectivity among neighboring voxels. Next, the inconsistency between the fuzzy memberships and the sampled and smoothed prior probability maps are calculated.

In 2009, Jeny Rajan, K.Kannan et al., [7] in their work based on FCD lesion analysis with Complex Diffusion Approach, the reason for selecting non-linear complex diffusion is that intra region smoothing will occur before inter region smoothing. So FCD and non-FCD areas in gray matter will defuse separately. The contrast between FCD areas and non-FCD areas will increase in the real plane after complex diffusion. The imaginary part of complex diffusion is almost equal to Laplacian of Gaussian (LOG), in which the borders will be highlighted. When the real part of the complex diffusion is divided with imaginary part, all the smooth areas in the gray matter will also get enhanced.

C. IMAGE SEGMENTATION

Image segmentation plays a crucial role in many medical imaging applications by automating or facilitating the delineation of anatomical structures and other regions of interest. Segmenting the structures or objects in an image is of great importance in a variety of applications including medical image processing, computer vision and pattern recognition. Different methods are applied to cortical tissues: WM, GM, and CSF.

In 1995, Simon Warfield, Joachim Dengler, Joachim Zaers, Charles R.G. Guttmann et al [14], they proposed the Automatic Identification of Grey Matter Structures from MRI to improve the Segmentation of White Matter Lesions, they developed a new algorithm for the development of the cortex. They have developed a segmentation method that uses the positive features of both statistical classification and

elastic matching methods to overcome the limitations. Elastic matching provides robust and accurate localization of these structures. This allows for improved segmentation of white matter lesions. A parzen window classifier is used to segment the volume into brain and non-brain classes. Intensity-based statistical classification and intensity in homogeneity correction are calculated simultaneously using the Expectation-Maximization (EM) segmentation algorithm.

In 2004, Faguo Yang, Tianzi Jiang, Wanlin Zhu, and Frithjof Kruggel [15], on their work based on developed novel and effective white matter lesion segmentation algorithm from volumetric MR images, their method is based on T1 and T2 image volumes. Firstly, we analyze those T1 slices, which have corresponding T2 slices. The segmented lesions in these slices provide location, shape and intensity statistical information for processing other neighboring T1 slices without corresponding T2 slices. This prior information is used to initialize a discrete contour model in the segmentation of the remaining T1-weighted slices.

In 2005, Jing Yang, Hemant D. Tagare, Lawrence H. Staib, James S. Duncan et al [16]., they proposed a level set based deformable model for the segmentation of multiple objects from 3D medical images using shape prior constraints. Their approach to multiple objects segmentation is based on a MAP estimation framework using level set based prior information of the objects in the image. We evaluate this level set distribution model by comparing it with the traditional point distribution model [4] using the Chi-square test. For our experiments, the mean distances show improvement in all these cases comparing with/without the level set based prior: average left and right ventricles, sub-cortical structures, amygdala and hippocampus.

In 2006, O. Colliot, PhD; T. Mansi, MSc; N. Bernasconi, MD, PhD et al [6], this paper presents a method for segmenting FCD lesions on T1-weighted MRI, based on two successive deformable models. The first deformable model is driven by feature maps representing known characteristics of FCD and aims at separating lesions from healthy tissues. The second evolution step expands the result of the first stage towards the underlying and overlying cortical boundaries, throughout the whole cortical section, in order to better cover the full extent of the lesion.

In 2007, Elsa D. Angelini, Ting Song, Brett D.Mensh, and Andrew F. Laine [17] presents Brain MRI Segmentation with Multiphase Minimal Partitioning: A Comparative study, the four segmentation methods that were applied to ten brains T1-weighted MRI for segmentation of cortical tissues: white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Segmentation errors are reported with comparison to manual labeling. The segmentation methods are intensity thresholding, fuzzy connectedness, Hidden Markov random field-expectation Maximization, and Multiphase three-dimensional level set. Addressing the in homogeneity issue, all four segmentation methods tested and perform a partitioning of the volumetric data into three tissue classes and a background relying on a strong assumption of tissue homogeneity for WM, GM, and CSF. Comparison to three other segmentation methods was performed with individual assessment of segmentation performance, statistical comparison of the performance, and evaluation of the statistical difference between the methods.

In 2008, Jacobus F. A. Jansen, PhD, Marielle C. G. Vlooswijk, MD et al[18], proposed on White Matter Lesions with Localization-Related Epilepsy, the performance of an automated WML detection algorithm, based on intensity thresholding, , a WML volume is calculated by collecting the hyper intense voxels after counting the number of voxels exceeding a predefined threshold of intensity, and K-Nearest Neighbor classification to segment GM, CSF, and WM, artificial neural networks, and fuzzy connected algorithms. WML were segmented from normal tissue by defining a global cut-off threshold on the images. These methods use only a single global intensity threshold to segment the WML for the whole brain or for each slice of the brain images.

D. Feature Extraction

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called *feature extraction.* To detect the lesion, GM, WM, and hyperintensity signal were extracted from MR images. The recent research works based on combination of different feature extraction and classification tools.

In 2003, Mohammad-Reza Siadat, Hamid Soltanian-Zadeh et al [19] presents the development of a human brain multi-modality database for surgical candidacy determination in temporal lobe epilepsy. The focus of the paper is on content-based image management, navigation and retrieval. The visual feature extraction module includes a set of applications each of which calculates a visual feature (e.g., color, texture, shape, and spatial relationship) within the segmented model and in a proper image modality. There are a variety of features such as volume, surface area, intensity mean-value and standard deviation, length, width, and principal vectors that are often of interest. These features are calculated once the segmented model is built. Using the extracted features, the classification module decides if the image set is going to be retrieved (on-line procedure). The result of classification is sent to the query module for further analysis and display to user. The clustering module performs the procedure of unsupervised indexing based on a portion of the extracted features.

In 2003, Marius George Linguraru, Miguel Ángel González Ballester, Nicholas Ayache [20] they presented a method of feature extraction for brain morphological studies. Using phase congruency, the detection results are not sensitive to image intensity and overcome common difficulties in brain imaging, such as the presence of a bias field. The method outperforms thresholding and gradient-based segmentation approaches and provides a good localization of features. Future applications of the method will focus on the detection of evolving tumors and multiple sclerosis lesions from temporal sequences of MR images. Sulci will be detected as structures with minimal temporal variations, in order to remove false positives.

In 2003, R. Tetzlaf, C. Niederhofer, P. Fischer [21], proposed the bioelectrical activity of a human brain in epilepsy would be analyzed using a Cellular Neural Network - Universal Machine (CNN-UM) proposed by Roska. Therefore a feature extraction method based on binary input-output patterns and Boolean CNN with linear weight functions called pattern detection algorithm is used. The treatment is focused on two

types of pattern occurrence that are defined as follows: **1.** A binary pattem occurs only once before a seizure and never occurs in any other recording. ***2.*** A binary pattern occurs frequently in all recordings of brain electrical activity never exceeding a maximum distance of ***N*** data segments between two occurrences. This distance is much smaller than the distance between the last occurrence of the pattern and the seizure onset.

In 2008, Madhubanti Maitra, Amitava Chatterjee, and Fumitoshi Matsuno [22] their present work proposed a method that uses an improved version of orthogonal discrete wavelet transform (DWT) for feature extraction, called Slantlet Transform, which can especially be useful to provide superior time localization with simultaneous achievement of shorter supports for the filters. The feature extraction from MR brain images can be carried out utilizing several popular signal/image analysis methods already available, e.g. independent component analysis, Fourier transform based techniques, wavelet transform based techniques etc. The discrete wavelet transform (DWT) is particularly useful for signal/image processing in the fields of de-noising, compression, estimation etc. An excellent classification ratio of 100% could be achieved for a set of benchmark MR brain images, which is significantly better than the results reported in a recent research work employing combination of different feature extraction and classification tools e.g. Wavelet Transform, Neural Networks and SVM.

In 2008, *Felipe P.G. Bergo, Alexandre X. Falcao* et al [23] have proposed the FCD segmentation using texture asymmetry of MR-T1 images of the brain. Their method works on volumetric MR-T1 images interpolated to an isotropic voxel size of 1.0$mm3$, and comprises the feature extraction, for each voxel p within the brain; we extract a 16×16 planar texture patch T_1 (p) tangent to the brain's curvature (as computed by the CR) and centered at p. The gradient vector of the CR distance transform at the voxel's location provides the surface normal. We also extract a symmetric patch T_2 (p), located at the reflection of T_1 (p) by the MSP. The patch size was chosen

experimentally. Smaller patch sizes did not provide good classification results, while larger patch sizes led to similar results with higher computational cost. For each patch we compute 6 features: sharpness (h), entropy, homogeneity, contrast, intensity mean (μ) and intensity standard deviation (σ). All features are scaled to fit within the [0, 1] interval.

E. Detection of FCD

FCD detection, a challenging and clinically valuable task that has not been addressed previously. We have to include the features from morphometric characteristics to the small lesions. While many techniques are being developed to detect FCD lesions from MR images. In most of the methods thickness map along with gradient techniques are used to compute FCD areas. The proposed method discusses the present conditions of the detection of FCD.

In 2002, Montenegro M.A, Li LM, Guerreiro MM, Guerreiro CA, Cendes F. [24], their work is based on FCD: Improving Diagnosis and Localization with Magnetic Resonance Imaging Multiplanar and Curvilinear Reconstruction, The diagnosis of FCD was based on the neuroimaging findings after a three step evaluation, always in the same order: (a) plain MRI films, (b) MPR, and (c) CR. For data analysis, we first assessed the contribution of the additional findings of MPR analysis compared with the results of the evaluation using only plain. MRI films, as is usually done in

routine practice. Second, we assessed the contribution of CR to the findings of plain.

In 2003, S. B. Antel1, N. Bernasconi, L. D. Collins et al [25], their work is based on an automated classifier to identify focal cortical dysplasia in patients with epilepsy was developed. The classifier was trained on 3D maps of first-order statistical and morphological models based on MRI characteristics of focal cortical dysplasia and 3D second-order maps constructed from second order texture analysis. A Bayesian classifier was trained on the maps of the first-order statistical and morphological models and three second order texture features to classify voxels within a T1 volume as CSF, GM, WM, GM/WM interface, GM/CSF interface, or lesional. The results of the classifier were compared to standard visual evaluation of presurgical MRI. Finally, they conclude strength of the classifier is its consideration of first- and second-order information from the T1-weighted MRI volume.

In 2006, O. Colliot_, T. Mansi, N. Bernasconi, V. Naessens et al [6], their work is based on a level set driven by MR features of focal cortical dysplasia for lesion segmentation. A method to segment FCD lesions on T1-weighted MRI, based on a 3D deformable model, implemented using the level set framework. Three MRI features drive the deformable model: cortical thickness, relative intensity and gradient. These features correspond to the visual characteristics of FCD and allow differentiating lesions from normal tissues. The proposed method was tested on 18 patients with FCD and its performance was quantitatively evaluated by comparison with the manual tracings of two trained raters. The validation showed that the similarity between the level set segmentation and the manual labels is similar to the agreement between the two human raters. This new approach may become a useful tool for the presurgical evaluation of patients with intractable epilepsy.

In 2006, Olivier Colliot, Samson B. Antel, Veronique B. Naessens et al [26], have proposed FCD on high-resolution MRI with computational models, On MRI, focal cortical dysplasia (FCD) is characterized by a combination of increased cortical thickness, hyper intense signal within the dysplastic lesion, and blurred transition between gray and white matter (GM–WM). Their methods are a set of voxel-wise operators was applied to high resolution 3D T1-weighted MRI in 23 patients with histological proven FCD and 39 healthy controls, creating maps of GM thickness, maps of relative intensity highlighting areas with hyper intense signal, and maps of gradient magnitude modeling the GM–WM transition Moreover, in all patients, the FCD lesion had at least two of these three characteristics.

In 2008, Christian Loyek, Friedrich G. Woermann and Tim W. Nattkemper [27], their work based on detection of FCD lesions in MRI using textural features, Focal Cortical Dysplasia is a frequent cause of medically refractory partial epilepsy. The visual identification of FCD lesions on magnetic resonance images (MRI) is a challenging task in standard radiological analysis. Quantitative image analysis, which tries to assist in the diagnosis of FCD lesions, is an active field of research. In this work we investigate the potential of different texture features, in order to explore to what extent they are suitable for detecting lesional tissue. The

results can show first promising results based on segmentation and texture classification.

III. COMPARISIONS AND DISCUSSIONS

In 1999, Yun Jang applied Gaussian distribution random number generator method for segmenting the MRI brain images and resulted with the detection rate of 77%. In 2003, Andrea Bernasconi proposed the MRI analysis methods for detection of FCD using the absolute gradient of gray level intensity based enhancement produces 87.5% detection rate. In 2006, O. Calliot got 75% detection rate by presented a method of preprocessing for intensity non-uniformity, intensity standardization and feature-based deformable model is used for segmentation of FCD lesions on MRI using level set evolution. In the same year, they proposed the Brain Extraction Tool for classifying the brain tissue by using histogram method, feature-based deformable model for segmenting the brain tissue, measuring the MRI image cortical thickness and relative intensity gradient value and finally, they got 70% detection rate. Again the same year, T.Mansi proposed preprocessing methods for intensity non-uniformity, intensity standardization and Gradient Vector Flow, automated histogram based segmentation methods which produced 75% accuracy. In 2007, Elsa D. Angelini proposed brain MRI segmentation with multiphase minimal partitioning: A comparative study. They got 86.7% detection rate by applying the segmentation methods are intensity threshold, fuzzy connectedness, Hidden Markov random field-expectation Maximization, and Multiphase three-dimensional level set. In 2008, Madhubanti Maitra proposed an improved version of orthogonal discrete wavelet transform (DWT) for feature extraction and results reported as 97% detection rate. In the same year, Felipe P.G. Bergo proposed the methods are intensity standard deviation, intensity mean, and gradient vector for detecting the FCD in MRI brain images which produced 94% detection rate. In 2009, Rajeshwaran Logeswaran got 80% by applied the dynamic histogram analysis for preprocessing and identified the brain tissue for segmentation.

IV. CONCLUSION

In this paper, the various techniques for detection of FCD is studied and presented. Also, a table is presented to summarize the previous research methods and their results studied. Using 3D MRI brain images for the detection of FCD, the maximum detection rate is 98%. In this case the FCD was identified by applying Surface-Based segmentation and blurred WM/GM segmentation using threshold classifier. Next 97% of detection rate was obtained by applying Fourier transform based feature extraction techniques, wavelet transform based feature extraction technique. For 3D MRI brain images the FCD detection rate is increased by combining both texture and morphological analysis.

REFERENCES

[1] O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi. “Segmentation of focal cortical dysplasia lesions on MRI using level set evolution”, Neuro Image Volume 32, Issue 4, 1 October 2006, Pages 1621-1630.

[2] Niels K.Focke, Mark R.Symms, Jane L.Burdett, and John S.Duncan, “Voxel-based analysis of whole brain FLAIR at 3T detects focal cortical Dysplasia”, Epilepsia, 49(5): 786-793.

[3] Leonardo Bonilha, Maria Augusta Montenegro, Chris Rorden et al., “Voxel-based morphometry reveals excess Gray Matter Concentration in patients with FCD”, Epilepsia 47(5):908-915,Blackwell Publishing, Inc 2006.

[4] Jan Kassubek, Hans-Jurgen Huppertz, Joachim Spreer, and Andreas Schulze-Bonhage, “Focal Cortical Dysplasia by Voxel-based 3-D MRI analysis”, Epilepsia 43(6): 596-602.

[5] Andy Khai Siang Eow.,”Quantitative Multi-modal Analysis of Pediatric Focal Epilepsy”, Massachusetts Institute of Technology.

[6] O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi, “Segmentation of focal cortical dysplasia lesions on MRI using level set evolution”, Neuro Image Volume 32, Issue 4, 1 October 2006, Pages 1621-1630

[7] Jeny Rajan, K.Kannan, C. Kesavadas, Bejoy Thomas, A.K. Gupta, “Focal Cortical Dysplasia (FCD) Lesion Analysis with Complex Diffusion Approach”, Volume33, Issue7, Pages:553-558.

[8] Rajeshwaran Logeswaran., “Computer Aided Medical image analysis for intra-operative Low-Field MRI in neurosurgery”.

[9] April Khademi, Anastasios Venetsanopoulos, Alan Moody,”Automatic Contrast Enhancement of WM lesions in FLAIR MRI”, Biomedical Imaging, From Nano to Macro 2009, IEEE International Symposium on Biomedical Imaging 2009. On page(s): 322 - 325

[10] Jun Yang; Sung-Cheng Huang, “Method for Evaluation of Different MRI Segmentation Approaches”, Nuclear Science Symposium, 1998. Conference Record.1998, IEEE, on page(s): 2053 - 2059 vol.3

[11] Andrea Bernasconi et al, “Advanced MRI analysis methods for detection of focal cortical dysplasia”, Epileptic Disorders. Volume 5, Number 2, 81-4, June 2003.

[12] Pierre Besson, Olivier Colliot, Alan Evans et al., “automatic detection of subtle FCD using surface-based features on MRI”, Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008, 5th IEEE international symposium on 2008, on page(s): 1633 – 1636.

[13] Shan Shen, Andre J. Szameitat, and Annette Sterr, “Detection of Infarct Lesions from Single MRI Modality Using Inconsistency Between Voxel Intensity and Spatial Location—A 3-D Automatic Approach”, Information Technology in Biomedicine, IEEE Transactions on 2008, Volume: 12 Issue: 4 On page(s): 532 – 540.

[14] Simon Warfield, Joachim Dengler, Joachim Zaers, Charles R.G. Guttmann et al [], “Automatic Identification of Grey Matter Structures from MRI to Improve the Segmentation of White Matter Lesions”, Journal of Image Guided Surgery, Volume1, Issuee:6, Pages: 326-338

[15] Faguo Yang, Tianzi Jiang, Wanlin Zhu, and Frithjof Kruggel, “White Matter Lesion Segmentation from Volumetric MR Images” Volume 3150/2004, Pages: 113-120.

[16] Jing Yang, Hemant D. Tagare, Lawrence H. Staib, James S. Duncan, "Segmentation of 3D deformable objects with level set based prior models", Proceedings IEEE international symposium on Biomed Imaging 2004 Apr 15; 1:85-88.

[17] Elsa D. Angelini, Ting Song, Brett D.Mensh, and Andrew F. Laine, "Brain MRI Segmentation with Multiphase Minimal Partitioning: A Comparative Study" International Journal of Biomedical Imaging, Volume 2007, Article ID 10526, 15 pages.

[18] Jacobus F. A. Jansen, PhD, Marielle C. G. Vlooswijk, MD, H. J. Marian Majoie, MD, PhD,et al[], "White Matter Lesions in Patients With Localization-Related Epilepsy", Invest Radiol 2008 Aug; 43(8):552-8.

[19] Mohammad-Reza Siadat, Hamid Soltanian-Zadeh, Farshad Fotouhi1, Kost Elisevich et al, "Multimodality medical image database for temporal lobe epilepsy", Proceedings Volume 5003, Medical Imaging 2003: PP.487-498.

[20] Marius George Linguraru, Miguel Ángel González Ballester, Nicholas Ayache, "A Multiscale Feature Detector for Morphological Analysis of the Brain", Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003, Volume 2879/2003, 738-745.

[21] R. Tetzlaf, C. Niederhofer, P. Fischer, "Feature Extraction in Epilepsy using a Cellular Neural Network based device-first results" Proceedings of the 2003 International Symposium on Circuits and Systems, Volume: 3, Page(s): III-850 - III-853 vol.3

[22] Madhubanti Maitra, Amitava Chatterjee, and Fumitoshi Matsuno, "A Novel Scheme for Feature Extraction and Classification of Magnetic Resonance Brain Images Based on Slantlet Transform and Support Vector Machine", SICE Annual Conference 2008, Page(s): 1130 – 1134.

[23] Felipe P.G. Bergo, Alexandre X. Falcao et al, "FCD segmentation using texture asymmetry of MR-T1 images of the brain", Biomedical Imaging, pages 424-427, IEEE, 2008.

[24] Montenegro M.A, Li LM, Guerreiro MM, Guerreiro CA, Cendes F, "Focal Cortical Dysplasia: Improving Diagnosis and Localization with Magnetic Resonance Imaging Multiplanar and Curvilinear Reconstruction", Neuroimaging 2002; 12(3): 224-230

[25] S. B. Antel1, N. Bernasconi, L. D. Collins et al, "Automated Detection of Focal Cortical Dysplasia based on Textural, Statistical and Morphological Analysis of MRI", Neuro Image, Volume19, Issue4, August 2003, Pages 1748-1759.

[26] O.Colliot, Samson B. Antel, Veronique B. Naessens et al [] ,"In Vivo Profiling of Focal Cortical Dysplasia on High-resolution MRI with Computational Models", Epilepsia, 47(1):134–142, 2006 International League Against Epilepsy.

[27] Christian Loyek, Friedrich G. Woermann and Tim W. Nattkemper, "Detection of Focal Cortical Dysplasia Lesions in MRI Using Textural Features ", Medizin 2008, Part 21, 432-436.

BIOGRAPHY

Dr. P. Subashini, Associate Professor, Dept. of Computer Science, Avinashilingam Deemed University have 18 years of teaching and research experience. Her research has spanned a large number of disciplines like Image analysis, Pattern recognition,

S.Jansi, have one year working experience as a Technical Assistant in Aeronautical Development Agency. Currently she is perceiving PhD in Image Processing. Area of Specialization: Image Processing, Neural Networks.

Journal of Computer Applications (JCA)
ISSN: 0974-1925, Volume IV, Issue 1, 2011

Search Engines:A Study

[1] Mr.K. Tarakeswar , [2] Ms. D. Kavitha

***Abstract*— The Internet is a huge collection of data. To get the appropriate information from it, using a search engine is the most effective way. Many Search Engines were introduced since 1990. In this paper we present a brief study on search engines. First, we present the definition of search engine, types of search engines and the general working process of a search engine. Then we give an example for the working process with a description of the Google search engine architecture. Later, we present a short description of the next generation search engines. Then we present comparisons among some major search engines.**

***Index Terms*— Internet; Search Engine; working; architecture.**

I. INTRODUCTION

Access to various types of information is necessary these days. The World Wide Web (WWW) contains a lot of web pages. To search for the information necessary for us from that huge collection of web pages, using a Search Engine will provide with efficient results. Many web pages in the WWW contain inappropriate information. This is due to the inappropriate naming and unnecessary highlighting of the content of web pages by their web masters. This raises the need of a search engine. Using a good Search Engine will filter out the necessary and relevant information needed by the user.

This paper presents an overview on the search engines. In the second section of this paper, we present the definition of search engine and we describe the types of search engines in the third section. We describe the general working of a search engine in the fourth section and present an example for it in the fifth section by explaining the Google search engine architecture. In the sixth section we present a brief explanation on the next generation search engines and in the seventh section we present comparisons among some major search engines.

II. DEFINITION OF SEARCH ENGINE

Definition 1: Search Engine is a program which searches the database, gathers and reports the information which contains the specified or related terms.

Definition 2: The term Search Engine [11] refers to the process of searching files using the key words specified. The key words found are returned and collated into the user information.

III.TYPES OF SEARCH ENGINES

Search Engines are of four types[6]. They are

A.Crawler based search engines.
B.Human powered directories.
C.Hybrid search engines.
D.Meta search engines.

A. Crawler based Search Engines

Crawler based search engines contain three parts. The first part is the 'Crawler' (bot or robot or spider). It is used to wander the web and create listings of web pages. The second part is the 'Index', which is a huge collection of copies of web pages and the third part is the 'Search Engine Software' which ranks the results. Because the crawler in this engine searches the web constantly, it provides updated information. Google, Live Search, Ask and most other search engines are crawler based.

B. Human Powered Directories

Human powered directories are search engines which depend on humans for their web page listings. These types of search engines get their listings of web pages from the submissions made by the respective web page masters. The submission contains the address, title and a brief description of the site. Later, the submission is reviewed by editors. A directory searches for results only from the page descriptions submitted to it. This is an advantage because, as the pages are submitted manually, the quality of the content will be better and more appropriate compared to the results retrieved by a crawler based search engine. But, the disadvantage is, any change made to an already submitted web page will not be updated until it is submitted again. Also, the ranking of pages can't be changed once ranking is done. Yahoo, dmoz and Galaxy are some examples.

C. Hybrid Search Engines

Hybrid search engines include the features of crawler based search engines and human powered directories. Currently, some search engines are using both features to provide effective results. MSN, Google and Yahoo are some examples.

D. Meta Search Engines

Meta search engines fetch results from other search engines. The fetched results are combined and ranked again according to their relevancy. These search engines were useful when each search engine had a significantly unique index and search engines were less savvy. Because the search has improved a lot, the need for these has reduced. MetaCrawler and MSN Search are some examples.

IV. WORKING OF SEARCH ENGINE

The working [3], [4], [5] of Search Engine involves three basic tasks. They are,

A.Searching the WWW and collecting the pages.
B.Keeping the index of the words they find and where they were found.
owing users to search for words or a combination of rom the index by using efficient software.

Manuscript received Apr 18, 2011.

Mr.K. Tarakeswar, Department of Computer Science and Engineering, G. Pulla Reddy Engineering College, Kurnool-518002, Andhra Pradesh, India. (e-mail : eshwartarak158@gmail.com)

D. Kavitha, Department of Computer Science and Engineering,G. Pulla Reddy Engineering College, Kurnool-518002, Andhra Pradesh, India. (e-mail : dwaramkavithareddy@gmail.com)

These tasks are performed by the three parts of a search engine. They are,

1) Crawler
2) Index
3) Search Engine Software.

The working of a search engine is shown in the Fig.1.

A. Searching theWWW and collecting the pages

Definition of Computer Robot, Spider or Crawler:

Computer Robots [10] are programs, which automate repetitive tasks at speeds impossible to be done by humans. The term 'bot' on the internet implies anything which interfaces with the user or collects data.

To present the result pages for a query a search engine must search and collect it. To find the web page from the millions of web pages present, search engines use the software robots called 'Crawlers or Spiders'. They build lists of words found in the web pages. This process of building lists is called 'Web Crawling'. A lot of pages must be traced to collect a useful list of words.

A spider chooses a list of heavily used servers and popular web pages as its starting point. It then begins with a popular web site, indexes the words present in it and also follows every other link present in that page. In this way, it quickly starts to travel spreading across widely used parts of the Internet.

The crawler carefully chooses at each step about which page to index. Some policies were introduced to guide the crawler. They are

1. Selection policy: Selection policy states which pages to download.
2. Revisit policy: Revisit policy states when to check for changes in web pages.
3. Politeness policy: Politeness policy states how to avoid overloading of web sites.
4. Parallelization policy: Parallelization policy states how to coordinate the different web crawlers distributed.

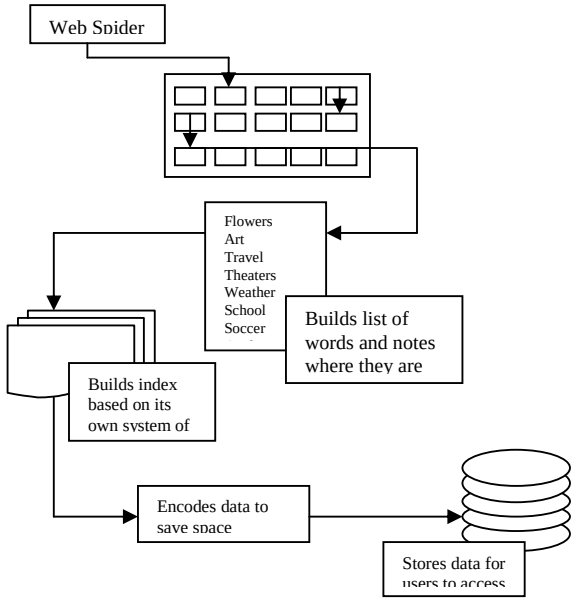


Figure1. Working of a Search Engine

B. Keeping an index of the words they find and where they were found.Before describing this task, we give a brief explanation on Meta Tags.

Definition of Meta Tags: Meta Tags[7] allow a web page owner to mention key words and concepts under which h page will be indexed. They guide a search engine in choosing appropriate meaning for a word from the several possible.

But, over reliance on meta tags leads to pages with popular topics, but which have very less or irrelevant content. To compensate this, the crawlers correlate the meta tags with the page content. They reject the meta tags which don't match with the words in the page.

After collecting the information, it must be stored in a way useful to the user. The stored data is encoded by the search engines to save the storage space. Two important components are present in making the collected data accessible to the user. They are mentioned below.

1. The information stored with the data

A search engine can store only the word and URL (Universal Resource Locator). This is a simple way of storage. In this case, the results cannot be ranked for their relevancy. To provide relevant results, weights can be assigned to the words based on their locations in the page.

2. The method by which the information is indexed

Indexing of words is made to allow the information to be accessed as fast as possible. An effective way is to use a Hash table for indexing. In the Hash table indexing we apply a formula for attaching a numerical value to the words. The formula used must evenly distribute all the entries.

In a dictionary, more pages will be present for the words starting with the letter's' than for the letter 'z'. So, the time to search for a word starting with's' is more, compared with the time taken to search for a word starting with'z'. Hashing evens out such differences. It also reduces the average search time for an entry. The hash table will contain the hashed values and pointers to the actual data. Hence, using efficient indexing and effective storage methods provide quick and better results for complicated queries also.

C. Providing results by using efficient search engine software

The third task is performed using search engine software. This software sifts through the results and ranks them according to their relevancy. Some basic principles are followed by all search engines to determine the relevancy of results. They are,

- Principle 1: The location of key words in a web page is a factor for determining the relevancy. The pages containing the search term in its HTML (Hyper Text Markup Language) tag, at the beginning of the page, in the links or subheadings and meta tags are more relevant.
- Principle 2: Frequency of key words in the page is another factor for determining the relevancy. The page with more occurrences of a search term is said to be more relevant.

Each search engine has its own method for assigning weights. Because of this, for the same query, different search engines provide differently ordered results.

1. Off Page Factors

Off Page Factors[10] are also used to rank web pages. They do not depend on the content of the page. They are

- Factor 1: Look of the web page.

Search engines infer a lot about the content of a page with a look of the page. Sophisticated techniques exist to find cial, fake and useless links and remove them.

- Factor 2: Click Through Measurement.

This determines the behavior of the user in relation to what results they choose while searching.

V. The Google Search Engine Architecture

In the Google search engine[1],[2], the three tasks of a search engine are performed as follows. The Google Architecture is shown in Fig. 2.

A. Searching the WWW and collecting the pages

The first task is performed using several distributed crawlers. The URL Server will send the lists of URLs to be fetched to the crawlers. The fetched web pages are sent to the Store Server. The Store Server compresses the web pages and stores them in a repository. Each web page is assigned a 'docID'. It is assigned each time a new url is parsed out of a page.

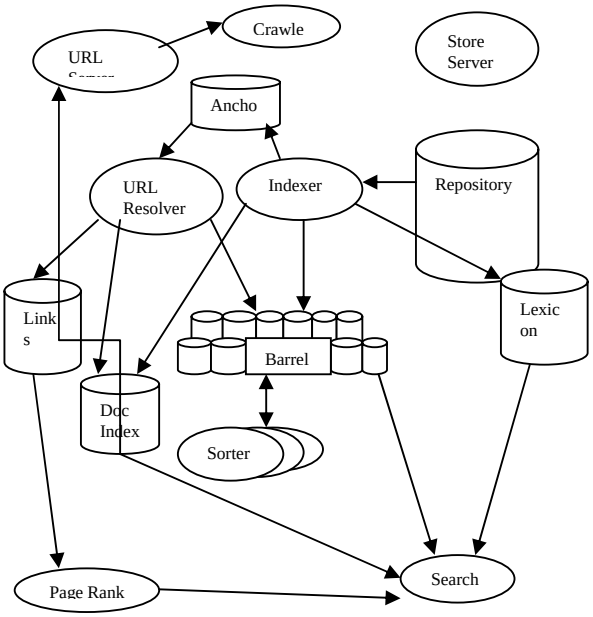


Figure.2 The High Level Google Architecture

B. Keeping an index of the words they find and where they were found.

In Google, Indexing is done by the Indexer and Sorter. The Indexer reads the repository to uncompress the documents. It then parses the documents. Every document is converted into a set of word occurrences which are referred to as 'Hits'. The Hits contains the words, their position in the document, their font size and capitalization. These hits are distributed by the Indexer into a set of Barrels, thus creating a partially sorted forward index. The Indexer also parses out the links in all web pages and stores the key information about them in 'Anchors' file.

The URL Resolver reads the Anchors file, converts relative URLs into absolute URLs which are then converted into docIDs. It also puts the anchor text into the forward index according to their docIDs. It generates a database of links which are used to compute page ranks of all documents. The Sorter takes the Barrels which are sorted by docIDs. These are resorted according to their wordIDs to create the inverted index. The Sorter produces a list of wordIDs and also offsets into the inverted index.

C. Providing results by using efficient search engine software.

The Dump Lexicon program takes the list generated by Sorter along with the lexicon generated by the Indexer. It then produces the lexicon which is used by the Searcher. The Searcher is run by a web server. It uses the lexicon built by the Dump Lexicon program, inverted index and page ranks to efficiently answer the queries.

VI. Next Generation Search Engines

The next generation search engines are referred to as Peer-to-Peer Search engines. They employ major types of discovery methods which are mentioned below.

- Selective forwarding systems.
- Flooding broadcast of queries.
- Centralized indexes and repositories.
- Decentralized hash table networks.
- Distributed indexes and repositories.
- Relevance driven network crawlers.

The Peer-to-Peer search implementation has two models. They are

A. Centralized server-client model.

B. Decentralized model.

A. Centralized server-client model.

The Centralized server-client model[9] contains a single, centralized server. It contains a directory of the shared files which are stored on the computers of users in the network. When a user searches for some file, the central server creates a list of files from its database of files which belong to users currently connected to the network. The server displays that list of files to the user. After the user chooses the file, a direct connection is setup with individual computers which contain that file at that moment. Opennap, kazaa and eDonkey are examples of Centralized server-client models.

Advantages

- The single, centralized index locates files quickly and efficiently.
- The search requests are sent to all clients who have logged in to the network. So, the search will be as through as possible.

Disadvantages

- The centralized server results in a single point of failure.
- As the centralized index is updated only periodically, the client may receive outdated information.

B. Decentralized model.

Decentralization of the network is made so that each peer can communicate as an equal to all the other peers. The Decentralized model[8] will not be having a single, central server. This model can be explained as follows.
Let there be some peers a, b, c, d, e, f etc., Whenever a peer 'a' enters the decentralized network, it connects to another peer 'b' to announce that it is alive. The peer 'b' announces to all other peers to which it is connected about the peer 'a'

being alive. The other peers c, d, e, f etc., repeat this pattern. After 'a' announces that it is alive, it can send search requests to 'b'. 'b' will pass this request to c, d, e, f etc., If 'c' has a copy of the file requested by 'a', 'c' sends a reply to 'b'. 'b' passes this reply back to 'a'. 'a' then opens a direct connection to 'c' and downloads the file. This scenario allows for an infinite network. In practice, a time to live (TTL) is used to limit the number of nodes reached by a request. Gnutella, mnet, freenet and gnunet are examples of Decentralized model.

Advantages

- The problem of a single point of failure is eliminated.
- The network is harder to shutdown.

Disadvantages

- Searching is slower in a decentralized network.
- Because of the TTL, the request for a file can't reach the node which will be having the file needed.

VII Comparison of Search Engines

In this section, we present comparisons among some major search engines based on some factors [12], which make a search engine provide satisfactory results. The results of the comparisons are presented in the below table, Table I.

TABLE I COMPARISONS OF MAJOR SEARCH ENGINES

	AltaVista	**Yahoo**	**Google**	**Ask**	**Teoma**	**MSN Search**	**Bing**
Links to a URL	No	Yes	Yes	No	No	No	No
Languages provided	All or English	41 languages	44 languages	6 languages	10 Languages	38 Languages	41 languages
Similar pages	No	No	Yes	No	No	No	No
Boolean and Phrase search	Yes	Yes	Yes	Yes	Yes	Yes	Yes
News and Multimedia search	Yes	Yes	Yes	Yes	No	Yes	Yes
Stemming	No	Yes	Yes	No	No	Yes	No
Other databases provided	Yes	Yes	Yes	Yes	No	Yes	Yes
Word in URL	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Search by File Type	Yes	Yes	Yes	Yes	No	Yes	Yes
Truncation	Yes	Yes	No	Yes	No	Yes	Yes
Grouping and Sorting Results	No	Yes	Yes	No	Only Grouping of results	Yes	Yes
Domain Search	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Thumbnails of results	No	No	Yes	Yes	No	No	Yes
Personalize	No	Yes	Yes	Yes	No	No	Yes
Date Limit Search	Yes	Yes	Yes	Yes	Yes	No	No

VIII CONCLUSIONS

Concluding this paper, using a Search Engine is obviously good to gather the necessary information. Many search engines have been developed to provide the best results for users. The present day search engines provide a variety of results like geographic search, domain search, personalization etc., but, they are unable to present satisfactory results for scientists, analysts, research students etc. To compensate this, the Peer-to-Peer search engines are being developed, which are referred to as the next generation search engines. The Peer-to-Peer search engines use the major searching techniques like flooding broadcast of queries, selective forwarding of queries, relevance driven network crawlers etc., They also use scalable and self-organizing algorithms and data structures and the results provided by them will be more quick and efficient compared to the present day search engines.

REFERENCES

[1] Sanjay Ghemawat, Howard Gobioff & Shun-Tak Leung, "The Google File System", Proc. The Nineteenth ACM Symposium on Operating Systems Principles, pp. 29-43, 2003.

[2] William Yip & Dr. Liz Quiroga, "Google Page Rank Algorithm", LIS 678 Personalized Information Delivery, Oct 11, 2008.

[3] Mark Levene, "An Introduction to Search Engines and Web Navigation", John Wiley & Sons, Inc., 2010.

[4] Fidel Cacheda, Diego Fernandez & Rafael Lopez, "Experiences on a Practical Course of Web Information Retrieval:Developing a Search Engine", Proc. Second International Workshop on Teaching and Learning of Information Retrieval, 2008.

[5] Curt Franklin, "How Internet Search Engines Work", [online] Available at: http://computer.howstuffworks.com/internet/basics/search-engine.htm

[6] J. M. Kassim & M. Rahmany, "Introduction to Semantic Search Engine", Proc. International Conference on Electrical Engineering and Informatics, 2009.

[7] Pegah Pishva & Mousa Majidi, "Study of HTML Meta-Tags Utilization in Web-based Open-Access Journals", Journal of Information Sciences and Technology, Vol. 22 Number 3(4-2007), 2007.

[8] Gabor Vincze, Zoltan Pap & Robert Horvath, "Peer-to-Peer based distributed file systems", International Journal of Internet Protocol Technology, Vol. 2, Number 2/2007, pp. 117-123, 2007.

[9] L. Plissonneau, J. L. Costeux & P. Brown, "Detailed Analysis of eDonkey transfers on ADSL", Proc. Second Conference on Next Generation Internet Design and Engineering, 2006.

[10] S. Brin & L. Page, "The Anatomy of a Large-scale Hypertextual Web Search Engine", Proc. the Seventh World Wide Web Conference. Brisbane, Australia.

[11] Wen-Jen Yu, Shrane Koung Chou, "A Bibliometric Study of Search Engine Literature in the SSCI Database", Journal of Software, Vol5, No 12 (2010), 1317-1322, Dec

[12] Ran Hock, (2010), "Major Search Engines – Features Guide", [online] Available at: http://extremesearcher.com/sechart.pdf

BIOGRAPHY

K. Tarakeswar obtained his B.Tech degree from Sri Krishna Devaraya University, Anantapur in the year 2009. He is pursuing his M.Tech in Computer Science and Engineering from Sri Krishna Devaraya University, Anantapur, India. He presented a survey paper at a national level conference.

D. Kavitha obtained her B.Tech degree from Sri Krishna Devaraya University, Anantapur and M.Tech degree from Jawaharlal Nehru Technological University, Anantapur, in the years 2001 and 2005 respectively. She is pursuing her Ph.D. from Sri Krishna Devaraya University, Anantapur, India. She is working as an Associate Professor in the Department of Computer Science and Engineering at G. Pulla Reddy Engineering College, Kurnool, Andhra Pradesh, India. She presented six research papers in international journals and five in national and international conferences so far. Her research areas include Computer Networks and Network Security.

K.S.R. COLLEGE OF ENGINEERING
Tiruchengode–637 215

JOURNAL OF COMPUTER APPLICATIONS

JOURNAL SUPSCRIPTION FORM

Yes, I would like to subscribe Journal of Computer Applications

Subscription Detail for printed copy

SINGLE	ANNUAL	THREE YEARS
Rs. 600	Rs. 2000	Rs. 5200
$ 40	$ 120	$ 325

From (Personal Name) :_____ Position :_____

Institution/ Organisation's Name :_____

Institution/ Organisation's Address :_____

Postal Code :_____

E-mail Address :_____ Website (url) :_____

Tel. :_____ Fax :_____ Mobile :_____

Please fill in this order form and mail to publisher Address

We are enclosing a D.D. for Rs. _____ in favour of " The Principal, K.S.R. College of Engineering payable at Tiruchengode".

D.D. No. :_____ Bank Name :_____ Date :_____

Signature_____ Date:

Approved by: _____ (Office) Date:

Publisher Address
Department of Computer Applications
K.S.R. College of Engineering
KSR Kalvi Nagar
Tiruchengode - 637 215
Namakkal District, Tamilnadu, India

Phone: 04288 - 274741 Ext. 570
Fax: 04288 - 274757
E-mail: ksrjca@gmail.com